74621

PROCEEDINGS

OF THE

TWENY-FIFTH ANNUAL FREQUENCY CONTROL SYMPOSIUM

Sponsored by

SOLID STATE AND FREQUENCY CONTROL
DIVISION
ELECTRONIC COMPONENTS LABORATORY
U. S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY

Details of illustrations in this document may be better studied on microfiche.

D D C
AUG 7 1972
C

[redacted]

ELECTRONIC INDUSTRIES ASSOCIATION
2001 Eye Street, N.W.
Washington, D.C. 20006

[redacted]

[redacted]

26 - 28 April, 1971

The Shelburne Hotel
Atlantic City, New Jersey

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Best
Available
Copy

ADDENDUM

to

PROCEEDINGS OF THE 25th ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

The following cross-reference relates to figures as mentioned in the paper by Messrs. E. A. Gerber and R. A. Sykes - A Quarter Century of Progress in the Theory and Development of Crystals for Frequency Control and Selection - to the figures as numbered and reproduced at the end of the text. Please insert this following Page 7.

LIST OF FIGURES

Fig. #	Reference in Text	
1	SLIDE	Number of papers given in fields of synthesis, fundamental studies of crystals and atomic & molecular frequency control.
2	SLIDE	Number of papers in various categories selected for review.
3	59-17-4;	Growth rate of synthetic quartz.
4	59-462-1;	Synthetic quartz-bar growth.
5	65-669-pg 672;	Irradiated sections of synthetic quartz.
6	65-5-2;	Piezoelectric materials under study.
7	57-62-6;	Loss peaks in natural and synthetic quartz.
8	58-84-7;	Loss peaks in irradiated quartz.
9	60-1-1;	Acoustic absorption at 5 MHz between -60° and 100°C in some representative samples of synthetic and natural quartz.
10	67-3-2;	Mode shapes in quartz plates.
11	67-3-5;	Computed and measured resonances for rectangular AT-cut quartz plates. Length c/thickness b=20.
12	67-3-7;	X-ray diffraction topographs of thickness-shear, thickness-twist and flexural modes in a rectangular AT-cut quartz plate $\frac{c}{b} = 29.8$.
13	66-1-2;	Topographs of modes in AT-cut quartz plates.
14	63-88-19;	Mechanical Q as a function of distance to diamond ground edge of wafer.
15	64-93-4;	Resonator response with rectangular electrode in X-direction, electrode length = 100b.
16	64-93-6;	Resonator response with rectangular electrode, length 10b.
17	66-252-2;	Dispersion curves for thickness shear and flexure in a quartz plate.
18	63-248-4;	Frequency change of precision crystal resonators due to temperature gradient.
19	61-22-6;	Frequency-temperature curves for AT, BT and RT-cut quartz resonators.
20	70-55-9;	Locus of turn-over temperature for AT-cuts.
21	61-49-1;	Relative frequency change per unit of applied radial force for AT-cut resonators.
22	57-463-3;	Simplified circuit of AN/TSM-15 Test set.
23	61-98-6;	Typical accuracy curve, 12 MHz 3rd overtone crystal unit.
24	64-444-1;	CI meters available or under development in 1964.
25	64-441-8;	Stabilization factor vs. frequency Crystal unit transmission test network.
27	68-259-8;	Block diagram of self tuned phase detection measuring system.
28	69-93-2;	Block diagram of measuring system using vector voltmeter.
29	69-93-5;	Resistance & drive level graph.
30	69-102-8;	Influence of design of π networks on phase errors.
31	63-316-1,2,3;	Test circuit for measurement of unwanted modes.
32	65-125-TabI;	Pressure rise leak-detection methods.
33	65-125-TabII;	Actual leak-test methods.
34	57-277-2;	Variation in Q with temperature and frequency due to intrinsic loss in quartz.
36	58-282-2;	Flexure mode crystal unit design.
37	59-423-1;	Parameter and cost ratios of crystal units.
38	60-200-1;	Oscillator for severe environments.
39	60-200-9;	Crystal unit design for high acceleration.
40	63-4-2;	Effects of oven and osc. shutdown.
41	63-4-3;	Effect of oven and osc. shutdown.
42	63-289-3;	Aging of metal enclosed crystal units.
43	63-289-4;	Aging of glass enclosed crystal units.
44	64-217-TabII;	Comparative failure rates of representative electronic components.
34	60-302-1;	Resistance change in solder and thermocompression bonded crystal units.
46	64-166-3;	Cold welded quartz crystal enclosure.
47	64-166-2;	Crystal unit standard enclosures.
48	64-166-13;	Aging of metal and glass enclosed 5 MHz crystal units.
49	66-192-3;	Frequency stabilization of precision 5 MHz crystal quartz units after oven shutdown.
50	68-89-10;	5 MHz glass enclosed precision crystal unit.

Fig. #	Reference in Text	
51	68-89-21;	Effects of oven and osc. shut-down.
52	68-89-20;	Aging characteristics of precision 5 MHz unit.
53	66-103-4;	Effects of SiO dielectric tuning.
54,55	70-126-13,14;	Use of ion etch techniques and resultant response characteristics.
56	68-67-3;	Typical 61 MHz crystal unit response showing trapped modes.
57	68-67-4;	Crystal resonator design algorithm.
58	68-67-15;	60 MHz response spectra using design to eliminate trapped modes.
59	64-487-1;	Compensated oscillator.
60	64-487-2;	Compnesation network.
61	64-487-18;	Critical areas of TXCO design.
62	65-617-2;	N-element compensation network.
63	65-617-16;	Results of successive compensation design.
64	69-192-pg 196;	Compensation design factors.
65	69-192-pg 197;	Characteristics of compensated CT-units.
66	61-297-10;	Frequency-temperature characteristics of the reference signal generator for AN/URC-32 transceiver.
67	61-297-11;	Frequency-temperature characteristics for SG-179(XA-3) reference frequency standard.
68	61-297-12;	Aging for two temperature compensation frequency standards.
69	70-191-3;	Frequency-temperature characteristic (computed) for a 6-point network.
70	70-157-7;	Power and transient characteristics of directly heated crystal plates.
71	56-339-4;	13 MHz crystal filter characteristics.
72	58-500-5;	Wide band crystal filer.
73	66-266-2;	Dual resonant crystal and equivalent circuits.
74	66-266-4;	Filter characteristics of single and tandem dual filters.
75	66-288-10;	Generalized lattice and ladder equivalent of multiresonator crystal filters.
76	66-288-12;	Normalized transmission characteristic of monolithic crystal filters.
77	66-309-6;	Equivalent circuit of dual resonator crystal.
78	66-309-10;	Attenuation characteristics of one section electromechanical filters.
79	67-179-12;	6-resonator monolithic crystal filter.
80	67-179-8;	X-ray topographs of critical frequencies of a 6-resonator filter.
81	68-188-9;	SSB narrow band HCM filter.
82	69-76-9;	Computed inband loss of monolithic filter.
83	70-84-10;	8-section monolithic filter; average and standard deviation of 24 MCF out-of-band characteristics.
84	70-84-11;	Average and standard deviation of 24 MCF in-band characteristics.
85	68-206-6;	Discriminator circuit and response of 3 resonator array.
86	68-206-9;	Impedance transformation of dual resonators.
87	68-206-10;	Low frequency generator using 3 resonator array.

TWENTY-FIFTH ANNUAL FREQUENCY CONTROL SYMPOSIUM

Sponsored by

U. S. ARMY ELECTRONICS COMMAND
Solid State and Frequency Control Division
Electronic Components Laboratory
Fort Monmouth, New Jersey

26-28 April 1971

The Shelburne Hotel
Atlantic City, New Jersey

EXECUTIVE COMMITTEE

General Chairman	Mr. V. J. Kublin
Technical Program Chairman	Dr. G. K. Guttwein
Executive Assistant	Mr. J. M. Stanley
Staff Consultant	Mr. M. F. Timm

TECHNICAL PROGRAM COMMITTEE

Dr. G. K. Guttwein
U. S. Army Electronics Command

Dr. E. Hafner
U. S. Army Electronics Command

Mr. M. Bernstein
U. S. Army Electronics Command

Mr. A. R. Chi
NASA, Goddard Space Flight Center

Mr. J. Dressner
U. S. Army Electronics Command

Mr. J. Giannotto
U. S. Army Electronics Command

Mr. D. L. Hammond
Hewlett-Packard Company

Dr. H. Hellwig
National Bureau of Standards

Mr. J. L. Jesperson
National Bureau of Standards

Mr. A. Kahan
U. S. Air Force

Mr. A. Machlin
U. S. Army Materiel Command

Mr. J. H. Sherman, Jr.
General Electric Company

Dr. W. J. Spencer
Bell Telephone Laboratories

Mr. R. R. Stone
U. S. Naval Research Laboratories

Dr. R. F. C. Vessot
Smithsonian Astrophysical Observatory

Dr. G. M. R. Winkler
U. S. Naval Observatory

HONORARY CRYSTAL INDSUTRY COMMITTEE

Mr. J. J. Colbert
Electronic Crystals Corporation

Mr. D. L. Hammond
Hewlett-Packard Company

Mr. J. D. Holmbeck
Northern Engineering Laboratories

Mr. J. H. Sherman, Jr.
General Electric Company

Mr. R. A. Sykes (Ret.)
Bell Telephone Laboratories

CHAIRMEN FOR TECHNICAL SESSIONS

RESONATOR THEORY

Dr. W. J. Spencer, Bell Telephone Laboratories, Inc.

APPLICATION OF FREQUENCY CONTROL DEVICES

Dr. W. J. Spencer, Bell Telephone Laboratories, Inc.

RESONATOR RESEARCH AND DEVELOPMENT

Mr. J. Wolfskill, Bliley Electric Company

Mr. Donald L. Hammond, Hewlett Packard Company

FREQUENCY AND TIME DISSEMINATION

Mr. J. Jesperson, National Bureau of Standards

Dr. G. Weiffenbach, Smithsonian Astrophysical Observatory

FREQUENCY GENERATION AND MEASUREMENT

Mr. Donald L. Hammond, Hewlett Packard Company

FILTERS

Dr. W. H. Horton, Piezo Technology, Inc.

ATOMIC AND MOLECULAR FREQUENCY CONTROL

Dr. L. S. Cutler, Hewlett Packard Company

INDEX

Page

Introductory Session

A Quarter Century of Progress in the Theory and Development of Crystals for Frequency Control and Selection
- E. A. Gerber, USAECOM
- R. A. Sykes, Bell Telephone Laboratories 1

History of Atomic and Molecular Control of Frequency and Time
- N. Ramsey, Harvard University..... 46

Resonator Theory

Acoustical and Optical Activity in Alpha Quartz
- R. D. Mindlin, Columbia University
- R. A. Toupin, IBM..... 58

Extensional, Flexural and Width-Shear Vibrations of Thin Rectangular Crystal Plates
- P. C. Y. Lee, Princeton University..... 63

Application of Frequency Control Devices

Quartz Crystal Applications in Digital Transmission
- R. B. Robrock II, Bell Telephone Laboratories, Inc..... 70

* Frequency Control Devices for Mobile Communications
- R. J. Nunamaker, Motorola, Inc..... 74

The Crystal Controlled Electronic Wrist Watch System: A Si-gate and CMOS-MSI Approach
- R. G. Daniels and F. H. Muca, Motorola, Inc..... 75

Frequency and Time in Airtraffic Control and Collision Avoidance Application
- V. I. Weihe, Consultant..... 82

Application of Crystal Clocks for Navigation and Time-Ordered Communication
- R. J. Kulpinski, MITRE Corporation..... 88

Time Synchronized Ranging System
- R. M. Aughey, Singer-General Precision, Inc.,Kearfott Div..... 94

Piezoelectric Sensors for Use as Pollution Detectors, Meteorology Monitors and Research Instruments
- J. Kertzman, Consultant..... 102

Resonator Research and Development

The Present State of the Art in Piezoelectric Sensors
- W. King, Esso Research and Engineering Company..... 104

Evaluation of Quartz for High Precision Resonators
- B. R Capone, A. Kahan, Air Force Cambridge Research Laboratory, and Baldwin Sawyer, Sawyer Research Products..... 109

* Abstract only - complete paper not received in time for publication

Page

Measurement of Vibration Modes of Piezoelectric Resonators by Means of Holography
- Y. Tsuzuki, Y. Hirose, and K. Ijiama, Yokahama National University of Japan..... 113

Study of Frequency Control Devices in the Scanning Electron Microscope
- R. J. Gerdes and C. E. Wagner, Georgia Institute of Technology 118

Quartz Crystal Units for High G Environment
- M. Bernstein, USAECOM..... 125

Using a Pendulum Diffractometer to Improve Precision of X-Raying Quartz Crystals
- G. E. Nemetz, General Electric Company..... 134

The Current Dependency of Crystal Unit Resistance at Low Drive Level
- S. Nonaka, T. Yuuki and K. Hara, Nippon Electric Company,Ltd., Japan..... 139

Standards and the Frequency Control Industry
- J. D. Holmbeck, Northern Engineering Laboratories..... 148

Frequency and Time Dissemination

Time Control of Frequency Shift Keyed Transmissions at VLF
- R. R. Stone, T. H. Gattis, Naval Research Laboratory, and T. N. Lieberman, NELEX..... 152

Omega VLF Timing
- E. R. Swanson, Naval Electronics Laboratory Center..... 159

Time Dissemination Capabilities Using the Omega System
- L. Fey, National Bureau of Standards..... 167

Use of Loran C Over Land
- B. Wieder, Institute for Telecommunication Services..... 171

One Way Time Dissemination from Low Altitude Satellites
- L. Reuger, Applied Physics Laboratory, Johns Hopkins University..... 179

Time Transfer by Defense Communications Satellite
- J. A. Murray, D. L. Pritt, L. W. Blocker, W. E. Leavitt, P. M. Hooton, Naval Research Laboratory, and W. D. Coring, Naval Electronics Systems Command..... 186

* Long Term Accuracy of Time Comparisons Via TV Radio Relay Links
- S. Leschiutta, Instituto Elet. Nazionale, Italy..... 194

Precision and Accuracy of Remote Synchronization Via Portable Clocks, Loran C, and Network Television Broadcasts
- D. W. Allan, D. D. Davis, B. E. Blair and H. E. Machlan, National Bureau of Standards..... 195

Methods of Local Time and Frequency Transfer
- D. H. Phillips, R. E. Phillips, J. A. Bowman, J. J. O'Neill, Naval Research Laboratory..... 209

International Coordinated Clock Time and the Coming Improvements in System "UTC"
- G. M. R. Winkler, U.S. Naval Observatory..... 217

Page

Frequency Generation and Measurement

Frequency Modulation Analysis with the Hadamard Variance
- R. A. Baugh, Hewlett-Packard Company..... 222

High Quality Quartz Crystal Oscillators: Frequency Domain and Time Domain Stability
- H. Brandenberger and F. Hadorn, Ebauches, S.A. Switzerland, D. Halford and J. H. Shoaf, National Bureau of Standards..... 226

The Design and Performance of an Ultra-Pure, VHF Frequency Synthesizer for Use in HF Receivers
- M. E. Peterson, Collins Radio Company..... 231

The Spectral Frequency Synthesizer
- D. R. Lohrmann and A. R. Sills, USAECOM..... 240

Filters

Energy Trapping in a Lithium Tantalate X-Cut Resonator
- K. Sawamoto, Elec. Comm. Lab., Nippon Tel. & Tel., Japan..... 246

CdS-Quartz Monolithic Filters for Use in the 100-500 MHz Frequency Range
- D. A. Roberts, Gould, Inc..... 251

Consideration About Channel Filters for a New Carrier Frequency System with Mechanical Filters
- H. Schussler, AEG-TELEFUNKEN, Germany..... 262

Semi-Monolithic Quartz Crystal Filters and Monolithic Quartz Filters
- L. Bidart, C.E.P.E., France..... 271

Monolithic Crystal Filters for Frequency Division Multiplex
- P. Lloyd, Bell Telephone Laboratories, Inc..... 280

Composite Filter Structures Incorporating Monolithic Crystal Filters and L-C Networks
- H. A. Simpson, E. D. Finch, Jr., R. K. Weeman, and A. N. Georgiades, Bell Telephone Laboratories, Inc..... 287

Atomic and Molecular Frequency Control

Cesium Atomic Beam Frequency Standards
- R. Beehler, NBS Boulder Laboratory..... 297

Frequency Biases in a Beam Tube Caused by Ramsey Cavity Phase Differences
- H. Hellwig, J. A. Barnes, D. J. Glaze, National Bureau of Standards..... 309

Performance of Newly Developed Cesium Beam Tubes and Standards
- R. Hyatt, D. Throne, L. S. Cutler, J. H. Holloway, and L. F. Mueller, Hewlett-Packard Company..... 313

A Field Independent Optically Pumped ^{85}Rb Maser Frequency Standard
- W. A. Stern, General Time Corporation, and R. Novick, Columbia University..... 325

Miniaturized, Rapid Warm Up Rubidium Frequency Source
- M. M. Zepler, Plessey Com., Ltd., England..... 331

Discussion of Cavity Pulling in Passive Frequency Standards
- J. Viennet, C. Audoin, M. Desaintfuscien, Lab de L'Horloge Atomique, France..... 337

Page

Hydrogen Maser Wall Shift Experiments at the National Research Council of Canada
- D. Morris, National Research Council, Canada..... 343

Hydrogen Maser Frequency Standard
- C. Finnie, R. Sydnor and A. Sward, Jet Propulsion Laboratory... 348

A QUARTER CENTURY OF PROGRESS IN THE THEORY AND DEVELOPMENT OF CRYSTALS FOR FREQUENCY CONTROL AND SELECTION

E. A. GERBER
W. Long Branch N.J.

and

R. A. SYKES
Bethlehem, Pa.

Summary

A brief history of the frequency control symposium is given together with a short report of the state-of-the-art in 1947. The progress and achievements in the past 25 years as presented at the various symposia will be reviewed in some detail covering the following categories: synthesis and production of piezoelectric materials, fundamental studies, crystal unit development and engineering, temperature control and compensation, filters and measurement and test. A brief concluding statement will be made as to the direction where future R&D efforts may lead us.

Introduction

The first symposium was held in a conference room of the Squier Laboratory in 1947. It was attended by personnel from the three services, contractors on the problems of frequency control, and members of the subpanel on frequency control of the old Research and Development Board. The purpose of this symposium was to review progress on the various contracts to assist the military in future program planning. During the next three meetings it was expanded to include others and went to Gibbs Hall. Due to larger attendance it moved to Asbury park in 1951, and stayed there until the first meeting in Atlantic City in 1960, the fourteenth symposium.

Let us take a look at the status in 1947 as we saw it then. Professor Cady's book on piezoelectricity had been published. Raymond Heising had published his book on advancements in theory, design and production during the war years. In addition, through the Galvin Industries and The Western Electric Company much of the manufacturing techniques had been made available to the industry. From a device standpoint plated high frequency crystal units had been developed which were stable under severe shock and vibration. This was a major improvement at that time. We had sealed metal holders and the dimensioning rules had been established for high frequency quartz plates. For precision crystal units such as used in the LORAN equipment, the 100 kHz GT had been developed. After three months of operation in the equipment, it was stable to about 1 part in 10^9 per day. In retrospect, by today's standards, in 1947 the holders leaked, high frequency crystal units aged or drifted badly and they had high resistance at low drive levels. There was considerable variation in resistance with temperature and many of them had unwanted modes which caused actual off frequency operation in the equipments.

One of the results of this study has been a discovery of the amazing amount of technical and other information that has been recorded in the proceedings of these symposia. Our advice to any young chap starting in the business today would be: go back and read the proceedings of past symposia. Make a note of the year and the article that is of particular interest and then when you meet a problem in the laboratory or in the development of equipment that rings a bell, re-read and study those articles. You will be surprised at the amount of time this will save and the amount of information you will find there.

To cover all of the developments in the past 25 years in the period assigned to us this morning would be an impossible task. However, we have picked a few categories which we think will be of interest to most of the people here.

General

Before we go into details of the review, it is interesting to have a more general look at the type of papers which have been presented during the years and to observe the ups and downs in technology manifested by the number of papers in the various technical categories. Slide 1 shows that our symposia started with a rather large effort in the synthesis of quartz and other crystals which were candidates for frequency control applications. As the growing of cultured quartz was more and more perfected, work on other crystals stopped as you can see from the declining numbers of papers in this category. Then, work in atomic and molecular frequency control gained momentum and it came in so strongly that a separate session was arranged in 1956. The papers given at this meeting are also counted in the overall number of papers in this field. In contrast to these two categories, the number of papers in the field of fundamental properties of crystals for instance, resemble more a quietly flowing river and they probably will continue to do so for some time to come.

Slide 2 shows a total of 92 articles of interest in the various categories shown that were selected from the proceedings of the Symposium published since 1956. Other authors might have selected differently and its quite possible not all our choices have been the best since it would require considerable study to do a proper selection. This slide serves only to show the areas that have received the most treatment throughout the years 1956-1970. Much of the published material shown here was of course, initiated in previous symposia of which the many contractual reports are the only source of information. For example, prior to 1956, the principles for the growth of quartz had been well established, the mathematical solution to the many vibration problems in quartz plates was well under way, the 2.5 and 5 MHz high frequency precision crystal unit designs were nearly complete, VHF crystals had been developed, initial work had been started on the use of thermistors for temperature compensation of crystal control oscillators, the impedance meters for measurement of frequency and resistance were under development and several papers had been published on manufacturing procedures. We will now proceed to give you a brief report of the progress in each category. No attempt will be made to establish credit but reference to

individuals or institutions may be made to more clearly identify the subject material.

Synthesis and Materials

We will start our more detailed review with the great success of having available today cultured quartz which is superior to natural quartz in many respects. Papers on this subject were presented by Bell and Brush-Clevite personnel at almost each symposium, joined later by Sawyer Research. Two growing methods finally yielded excellent results; the high pressure, (1200 - 1800 bar) faster method with NaOH solution used by Bell and Western Electric, and the low pressure, (600 - 900 bar) slower method with Na_2CO_3 solution used by Brush-Clevite-Sawyer. Success reached by all researchers was the result of very thorough, systematic investigations. Fig. 59-17-4* shows the results of a study of the growth rate on the basal plane as a function of the percent fill in NaOH. Also in 1959, success was reported in growing quartz bars with the length parallel to Y specifically suited for the fabrication AT cut plates as shown in Fig. 59-462-1. The composite nature of cultured quartz material is vividly revealed in Fig. 65-669-p. 672 which shows three irradiated sections cut from a Y - bar cultured quartz. The smoky sections are the +X and -X growth regions and they show slightly more internal friction than the uncolored Z - growth material. In both natural and cultured quartz, smoky color does not necessarily mean lowered Q. If a quartz plate is cut, with a +XZ boundary traversing its active region, larger losses may manifest themselves.

After a lull of ten years, new work on piezoelectric crystals other than quartz has been reported, specifically the five materials shown in Fig. 65-5-2 have been grown in crystalline form and their elastic, piezoelectric properties measured. Specifically, lithium niobate and lithium tantalate look very promising. While retaining the desirable features of quartz, such as high Q, hardness and good crystalline perfection, they are characterized by coupling factors which are about five times greater than those of quartz. These properties make them highly suitable for frequency control devices in the VHF range, especially in crystal filter applications.

Fundamental Studies

Work on quartz synthesis has awakened great interest in the fundamental properties of crystalline materials and this effort has led over the years to a profound knowledge of the lattice parameters and the influence of imperfections on Q and other resonator properties. A paper was given in 1957 which reported on the measurement of the resonance resistance of crystal units made from natural and synthetic quartz, between three and nine MHz and 1.2 to 400°K. For these and most of the later measurements, a precision 5th overtone glass-enclosed AT - cut vibrator has been used. This particular construction has been employed because its unique design suggests that most of the measurable loss is in the vibrator material and not in the mounting structure. Fig. 57-62-6 shows the resonance resistance of 5MHz crystal units. The location of the relaxation peaks was the same for all samples tested, but the magnitude varied greatly. Fig. 58-84-7 shows that the absorption peak at 50°K disappears after X-ray irradiation, but a peak at 100°K shown in this figure derives from substitutional Al. This peak is not apparent in quartz grown on Z-cut (Z-growth). Fig. 60-1-1 shows measurements of the acoustical absorption at higher temperatures. This rising acoustic absorption appears in thickness shear modes above room temperature and peaks at 130°C but occurs only in Z-growth quartz. This loss was first ascribed to the formation of sodium aluminum silicate but was shown at the 1967 Symposium to be correlated with OH in quartz. It was also reported at that time that lithium, used as a dopant of the growth solution, results in significant improvements in synthetic quartz. It approaches in its properties, including its frequency-temperature behavior, the best quality natural quartz. Other techniques have also been reported which can be used to eliminate the acoustic effects caused by the sodium-aluminum centers: X- or gamma ray irradiation, or electrolysis. The latter treatment has the added benefit of reducing the sensitivity of the resonance frequency to ionizing radiation.

Prof. Mindlin and his associates at Columbia University have made very important and lasting contributions on the mathematical theory of vibrations of crystal plates. Fig. 67-3-1 shows the various possible modes of vibration along X_1 and X_3 in a rotated Y - cut quartz plate. All these modes have overtones and they are coupled to varying degrees. These formidable vibration problems have been treated step by step, as reported at the various Symposia.

It is impossible to do justice to all of Prof. Mindlin's contributions, so we wish to focus your attention on what has been reported during the last few years. In the case of rectangular plates with all four edges free, solutions are not expressible in terms of a finite number of elementary functions. The additional complications of anisotropy and high frequency makes general solutions even more intractable. However a pruning of the general equations considering the specialities of the AT cut resulted in a set that could be solved in close form. This solution shown in a paper by Mindlin and Spencer contained an additional family of modes - the thickness twist overtones as shown in Fig. 67-3-2. The mode shape comprises phase reversals, across the width of the plate, resulting in a twisting deformation across the width. Fig. 67-3-5 shows a comparison of the computed and measured resonance frequencies of coupled thickness - shear and flexure and their twist overtones for a rectangular AT - cut plate. The agreement between theory and experiment is quite impressive. To identify more completely the measured resonances, X-ray diffraction topographs are shown in Fig. 67-3-7. The fundamental thickness shear mode ($n=0$, $m_2=1$) is not shown. Directly above this mode in frequency is the ($n=2$, $m_2=1$) twist overtone of the thickness shear mode for which a topograph is shown. The flexure components of displacement are still visible as banding of the shear displacement. The other 3 topographs show other twist overtones, one symmetric overtone of thickness shear and the 36th overtone of flexure.

The X-ray topographic method was reported the first time at the 1963 Symposium. Fig. 66-1-2 shows how powerful this method is. It is a frequency scan of a fundamental thickness shear AT resonator. X-ray topographs were taken at several resonance frequencies. The light sections dividing dark areas are nodes. Note the irregularities in the material and strained area due to the mounting structure.

Shockley and his coworkers of Clevite Corp. presented at the 1963 symposium a paper entitled "Energy Trapping and related studies of multiple electrode filter crystals". The conclusion of this important paper summarizes their results in a nutshell; therefore we quote it here verbatim. "Performance characteristics of high frequency quartz filter crystal units can be

*The figure numbers indicate the year of the symposium - the page on which the paper appears in the proceedings - and the actual figure number for that paper.

explained in terms of existing elastic wave theory and a cut-off frequency, which in effect confines the vibratory energy to a limited region surrounding the electroded portion of the wafer. Below the cut-off frequency, vibratory energy decays exponentially with distance away from the electroded portion of the wafer. This exponential decay is not associated with dissipation but rather acts to trap the vibratory energy within a confined region. This explains the high Q values observed for quartz resonators mounted on low Q supports and gives a theoretical basis for the small inter-resonator coupling observed on multi-resonator wafers. Application of the principles described here should lead to substantial improvement in Q of high frequency filter crystals and to further displacement in frequency of spurious responses from their principal resonance."

Fig. 63-88-19 shows the Q of a dot resonator as a function of the number of half wavelengths to the plate edge. The experimental values agree very well with theory which postulates the trapping of the oscillating energy for a certain lower value of d/λ. On the other hand, if most of the energy of the shear wave is confined to the plated portion, for example, the rate of decay into the unplated part of the crystal plate is rapid, then anharmonic modes due to the plate boundary will be subdued. The response spectrum of a 10 MHz AT cut resonator is shown in Fig. 64-93-4 where all anharmonic resonances below the cut-off frequency f_2 for the unplated part are trapped within the plated part of the plate. In Fig. 64-93-6, f_2 is much lower (the plating thickness is decreased) and very little trapping of anharmonic overtones appears. We will discuss the importance of these findings for filters later.

Mindlin's tremendous theoretical tools and experience came extremely handy in coming up with a theory for the phenomenon of "energy trapping." The dispersion curves for thickness shear and flexual waves in an infinite plate, propagated in the X -direction is shown in Fig. 66-257-2 which demonstrates the criticality of the electrode dimensions. Dimensionless frequency is plotted as ordinate and lateral wave number as abscissa. A detailed study of this graph will relate the electrode dimensions for trapped energy resonators. The upper curves are the dispersion curves for the uncoated and the lower ones for the coated part of the plate. To the left of zero, the wave number is imaginary and the waves are non-propagating. At frequencies between the two cut-off frequencies, the thickness - shear motion in the plated part is propagating, but in the unplated portion, we have again non-propagation so that the amplitude falls off exponentially outside of the plated portion. Above the cut-off frequency of the unplated portion, both waves are propagating over the entire plate. If the X - dimension of the plating is short enough so that the lateral wavelength of the first anharmonic overtone would have to be, roughly, two thirds the length of the plating, the wave number would be high and the frequency of the first anharmonic overtone would be above the cut-off frequency in the unplated portion. Then that first and all higher overtones would propagate out. The critical length of plating, then, will be inversely proportional to the wave number in the plated portion at the cut-off frequency of the unplated portion of the plate.

In 1956 a paper was presented on high precision crystals which were excited by electrodes creating a field parallel to the major surfaces of the plate. Since this arrangement left the most active center part of the plate free of coating, a higher Q resulted, bought, however, for the price of a higher impedance level of the crystal unit. The parallel - field excitation proved to be advantageous when the crystal was exposed to sudden temperature changes. Fig. 63-248-4 shows that the transient frequency change is one order of magnitude smaller than in the case of the perpendicular field.

Several excellent papers appeared during the history of our Symposium on precise determination of the temperature coefficient of the stiffness coefficient and the frequency. A paper given in 1961 investigated double rotated cuts and resulted in the introduction of a new cut. Its orientation according to IEEE nomenclature is $\theta = -34^{\circ}$; $\phi = 15^{\circ}$. Fig 61-22-6 shows that it has about the same minimum frequency deviation as the best AT cut, but displays a second order parabola. The main value of these studies, however, is the precise determination of first order temperature coefficients. Very high precision measurements on the same topic were presented at last year's Symposium. Fig. 70-55-9 shows the locus of the turnover temperature measured with 5MHz high precision AT resonators. The scatter of the measured points is seen to lie within a 5" band about the calculated curve, indicative of the angular error, and more important, of the homogenity of natural quartz.

Before we close our discussion on the more fundamental properties of crystals, we want to mention measurements of the influence of applied compressional forces on the frequency of an AT plate as a function of the azimuth as shown in Fig. 61-49-1. These measurements started a long series of similar measurements with many different types of forces, crystal sizes and orientations, reported at our Symposia after 1961. However, no final theoretical explanation for these effects is available today.

Measurement and Test

Papers in the field of measurement and test dealt mainly with two topics. The CI - Meter and the π - network. To a lesser extent, bridge circuits for measuring crystal parameters were discussed, and also methods for testing leaks in crystal enclosures. Fig. 57-463-3 shows the circuit diagram of the VHF crystal impedance meter AN/TSM - 15, reported at the 1957 Symposium. Compared with the older TS -683, it provides a circuit such that the crystal power dissipation is relatively independant of crystal resistance, important for production testing of crystal units. This was made possible by the low resistance values used in the crystal network and the addition of a grounded grid amplifier. A metering circuit to measure crystal voltage was also added. The TSM - 15 covers a frequency range from 75 to 200 MHz. The repeatability of frequency measurements, specifically in the HF range, had been a problem with CI - Meters, specifically with the TS - 330 which covers the range from 1 to 15 MHz. This problem was due to crystal current changes and was solved by the use of automatic gain control. This allowed control of power levels from one to ten microwatts. It resulted in an improvement of the frequency repeatability from 10^{-7} to 10^{-8}. As shown in Fig. 61-98-6 the twenty three data points have a scatter that rarely exceeds $5 \cdot 10^{-9}$.

A compilation of all CI - Meters available or under development in 1964 is shown in Fig. 64-441-1. The table is still valid today with the exception that the TSM - 20 is replaced by the TSM - 305 which is commercially available. Fig. 64-441-8 shows the advantage gained by an inductively tuned CI - Meter. We have a constant stabilization factor for each frequency band and at the same time, greatly reduced the number of necessary bands to cover the same overall frequency range. The problem with the low frequency TS-710 meter was the overdrive of good, low resistance crystals after rated drive was set as specified through a relatively high resistance. The solution was to improve the AGC and make the resistors in the crystal network adjustable for specified frequency bands.

Turning now to the π or transmission circuit method, the firm of Rohde and Schwarz presented in 1959 the first commercial version of the transmission circuit method,

together with synthesizers to cover the frequency range from 30 Hz to 30 MHz. Fig. 59-354-2 shows a side and top view of the transmission network. The two resistors R_{T1} and R_{T2} forming the network consist of four deposited - carbon resistors arranged in the form of a star. The two resistors connecting the network with the oscillator and the detector form the inner conductor of a coaxial line. Due to this construction, it can be used up to very high frequencies. A phase indicator connected to the crystal terminals facilitates tuning of the equipment to the crystal resonance frequency. The phase detection technique is shown in block diagram by Fig. 68-259-8. This forms a phase locked oscillator and allows continuous resonance frequency measurements for manufacture and test. A crystal measurement system shown in Fig. 69-93-2 using an RF vector voltmeter was described at the 1969 symposium. The vector voltmeter test set is basically an oscillating loop containing an amplifier, a phase shifter and the crystal in the π - network.

As in Fig 68-259-8, the phase angle across the crystal is sensed by the vector voltmeter and adjusted until the frequency of oscillation is of the proper value to produce a zero phase angle across the crystal. The vector voltmeter can be switched to measure the voltage to ground at either probe A or probe B. As shown in Fig. 69-93-5 these voltages can be used to find the crystal resistance and drive level. The voltage readings V_A and V_B determine the X and Y coordinates of a point which in turn determines a constant resistance line and a constant power curve. The basic accuracy of this crystal test set is about 3 to 5 parts in 10^8 and the resetability error in the order of 1 to 2 parts in 10^8. Crystal resistance and drive level accuracy is in the order of five percent.

Two papers, presented in 1969 and 1970, compared the precision of the π network with those of bridges. One author, comparing an admittance bridge and the GR bridge type 1609 with the π network, found no significant systematic differences. The frequency resetability appeared to be $\pm 7 \cdot 10^{-8}$ in the worst case, provided that the π - network was designed carefully and phase errors due to parasitic inductances and capacitances were compensated for. Fig. 69-102-8 shows how the phase error can be held small and independent of the calibrating resistance. On the other hand, according to the second paper, bridges seem to be better suited for the measurement of the motional capacitance above 10MHz.

Attention to the proper measurement of unwanted modes in crystals was drawn in a paper presented in 1963. The hybrid coil bridge method shown in Fig. 63-316-1,2,3 is suitable for this purpose if proper precautions are used. It shows the schematic of the test circuit and the equivalent circuit of a crystal vibrator. When the bridge is balanced the equivalent test circuit can be represented by the lower figure and the maximum insertion voltage ratio calculated.

Test of the seal and evaluation of the crystal unit enclosure is also an important step in developing and producing good crystal units. Methods for unit hermeticity fall into two categories; those which detect internal pressure rise and those which depend upon the detection of a trace substance which passes through an actual leak. Fig. 69-125-Tab I shows pressure rise detection methods together with their sensitivity. The ionization current method has been employed for some time, but is now being replaced. Among the actual leak test methods shown in Fig. 65-125-Tab II the bubble test is perhaps the easiest and least costly. Its sensitivity may be increased by subjecting the units to be tested to external helium pressure. The widely used mixed gas backfilled hellum mass spectrometer test can detect leaks of 10^{-10}cc/sec. It is excelled by the radioactive tracer technique, but the latter is economic only for the large volume production.

Crystal Unit Development & Engineering

Figure 57-277-2 shows the calculated variation in Q due to the intrinsic losses in quartz as a function of temperature and frequency. Both low and high frequency crystal units were developed for operation in the range of 4°K to 70°K. They had high Q and very low drift but continued effort in this direction ceased, due to practical difficulties of maintaining accurate control at these temperatures, and because more promising developments appeared possible for high frequency crystal units at 2.5 and 5 MHz at room temperature. Reduced drive levels and improved oscillator stability were obtained by using multi-stage gain as shown in Fig. 57-277-21.

Comparison of different types of flexure type crystal unit structures is shown in 58-282-2. The XY' flexure unit, while widely used in Europe for several years, had not received much attention in this country until presented here. It is shown that it has a lower second order temperature coefficient and increased stability with time over the other two designs. Within the frequency and impedance limitations it does have many advantages.

Fig. 59-423-1 gives the relationship between important characteristics of precision, moderate precision and standard crystal units prevailing at the time. As can be seen, for 100 times the stability one paid nearly 100 times the price. While the stability of all types has increased by today's standards the price differential has been materially reduced.

High frequency crystal units were developed to withstand high shock and vibration for use in missiles and satellites. The performance of production models at 5 MHz over a 10 hour aging period was less than 5×10^{-10}. The design of a complete oscillator and oven to withstand high shock and vibration is shown in Fig. 60-200-1. The variation in frequency due to vibration in the radial and longitudinal modes was less than $\pm 1 \times 10^{-9}$. An improved design of the crystal unit to withstand high acceleration rates is shown in Fig. 60-200-9. Note the biconvex plate design. The support ribbons were also preferentially located with reference to the crystal axes.

The next major step in reducing frequency drift with time resulted from a study of the on-off characteristics of precision oscillators. Fig. 63-4-2 shows the results of oven and oscillator shut down on the 2.5 Mhz oscillator. Crystal units processed to reduce the residual gases present in the enclosures resulted in the performance shown in Fig. 63-4-3. As can be seen the effects of oscillator shut-down are essentially nil and the effect of oven shut-down have been reduced. A strain relaxation of the mounting system or quartz material may account for the small change.

A study of many standard grade high frequency crystal units subjected to aging at high temperatures and at various levels of drive shows the results given in Fig. 63-289-3 for those units in solder sealed metal holders. It seems obvious that leaks had developed in most units. Those units sealed in glass holders gave results shown in Fig. 63-289-4. Except for one group of catastrophic failures the performance is good even at 500% drive levels. The failure rate of crystal units as a group, compared with other electronic components, is shown in Fig. 64-217-Tab II and is only surpassed to any great extent by mica capacitors.

It has long been known that the Q of low frequency wire mounted crystal units has been limited by the solder

connection. Fig. 60-302-1 shows the results of replacement of the solder connection by a thermo-compression bond. The improvement is evident in Q, particularly at the higher temperatures. It is obvious that losses in the mounting system even with thick crystal plates still limit the Q of these units.

An additional improvement in performance of crystal units came with the introduction of cold weld enclosures, the basic design of which is shown in Fig. 64-166-3. Fig. 64-166-2 shows a series of outlines including the frequency range covered using standard size crystal resonator elements. Similar cold weld enclosures have since been applied to the HC-6, HC-18, and flat-pack configurations. A summary of the drift characteristics of 5 MHz crystal units utilizing progressively improved enclosures and processing techniques is shown in Fig. 64-166-13.

Once the holder problem had been solved it was relatively easy to go the rest of the way and apply hydrogen firing of the piece parts and vacuum baking during low pressure cycles in an oil free system to reduce the contaminating atmosphere. Fig. 66-192-3 shows the results of oven and oscillator shut-down on crystal units with solder bond supports and thermocompression supports that could be high temperature processed. The indications are that the change in the solder units is due to sorption-desorption of residual gases and that shown for the high temperature bonded case is thermal relaxation. Crystal units shown in Fig. 68-89-10 are in production. Typical shut down characteristics of these units are shown in Fig. 68-89-21 and drift characteristics in Fig. 68-89-20. Thus, there seems little need for further development of these precision types to reduce drift since ultimate accuracy will be determined by atomic or molecular standards. Precision oscillators will maintain sufficient intermediate accuracy and future needs will require improvements in spectral purity. Improvement in the drift rate of general purpose type crystal units, however, is required.

While the concept of "energy trapping" is covered in the general subject of "Fundamental Studies" there are two interesting developments that have been reported. One is shown in Fig. 66-103-4 in which it is demonstrated that after sufficient electrode deposit has been made, a further reduction in frequency may be obtained by adding SiO to the entire plate surface without changing the conditions to trap additional modes. The other is shown on Fig. 70-126-13, where, by ion etch techniques, a pattern the size of the electrode has been etched into the quartz plate. This allows, in the case of VHF resonators, for an adequate thickness of metal to be deposited for an electrode to have low resistance before additional modes are trapped. Fig. 70-126-14 shows clearly the resonance spectrum for different degrees of plateback.

Throughout the Proceedings a large amount of design information and procedures can be found. For example, the response spectrum of a 5th overtone 61 MHz crystal unit given in Fig. 66-67-3 results if too large or too thick electrode is employed. Many of the anharmonic modes become trapped. One procedure outlined in Fig. 68-67-4 leads to the results shown in Fig. 68-67-15.

With the mathematical theory of vibrating plates well established, together with the trapped energy concept well understood, most of the conventional crystal unit designs may be obtained directly from the computer. As better design and more accurate constants become known the programs may be improved or altered to give design information to meet more sophisticated requirements.

Temperature Control and Compensation

A number of papers were presented on the development of ovens and oven control, particularly on proportional control. Also a great deal of work was carried out on change-of-state ovens during the period when low temperature operation was being considered. The fact that they were useful only at fixed temperatures and a high state of solution purity was necessary, led to abandonment of this approach. Since the temperature-compensation technique has received the most attention in the proceedings, most of this section will be devoted to it.

L. F. Koerner reported as early as 1955 on temperature compensation by thermistor shunted capacitors to reduce the frequency change of crystal units due to temperature. This method introduced additional loss to the crystal unit and was considered undesirable at the time. With the increased availability of stable thermistors and variable capacitance diodes, the former method has received little attention. Most of the work in this area has since been concentrated on the non-dissipative thermistor-varactor approach.

An oscillator circuit employing this principle is shown in Fig. 64-487-1. A reverse biased diode C_V is in series with the crystal unit. A change in bias will vary the effective capacitance of the diode with a resultant change in frequency of the oscillator. It therefore, is necessary to develop a voltage V_D as a function of temperature to just compensate for the normal frequency-temperature characteristic of the crystal unit. This gives rise to the term "Temperature Compensated Crystal Oscillator" or TXCO for short. The circuit shown in Fig. 64-487-2 employing resistors and thermistors will provide most of the voltage-temperature characteristic required for AT type crystal units. A single thermistor in the shunt arm may be used if only a linear negative coefficient is to be compensated. The principal problem is then, to determine the values of applied voltage, resistors and thermistors in the compensation network needed to reduce the deviation in frequency of the oscillator to the required level. Many factors must be considered. Fig. 64-487-18 lists those elements that must be considered in the four categories: Crystal unit parameters, variable capacitance diodes, compensation network and oscillator circuitry. With a greater number of elements in the compensation network the frequency change with temperature may be reduced materially. Such a network is shown in Fig. 65-617-2. By employing computer techniques, since curve fitting is a primary problem here, a suitable design may be developed quickly. Fig. 65-617-10 shows the results of successive runs to achieve reasonably good compensation. In any case, as the temperature range and frequency requirements are increaded, frequency compensation becomes more expensive. It has been suggested that for wide temperature requirements, an oven to limit the lower temperature to 0°C, for example, might result in a more economical solution. Also a proportional control oven to meet severe frequency requirements should be considered.

Fig. 69-197-pg 196 shows the detailed procedure for estimating the voltage required for compensation. By comparison of the inverse frequency-temperature characteristic of the crystal unit, the computed shift in frequency with series capacitance and the varactor characteristics it is possible to determine the required voltage that must be developed for compensation. This figure also shows the use of a d.c. amplifier to permit low voltages to be used, an offset voltage to take care of initial calibration as well as aging, and inductance associated with the varactor to make it more linear. An example of compensation for GT type crystal resonators in the frequency range of 800 to 1500 kHz is shown in Fig. 69-197-pg 197. Also shown is the change to be expected from a ± 20 p.p.m. adjustment of the frequency.

Nearly one order of improvement in frequency tolerence is shown.

As early as 1961 compensation to two orders of magnitude have been possible if sufficient complexity is allowed for the network and computer method used. Fig. 61-297-10 shows the results of development for the AN/URC-32 Tranceiver. Nearly as good results were achieved over an even greater temperature range for the SG-179A Reference Frequency Standard as shown in Fig. 61-297-11. The long term drift of these types is shown in 61-297-12 and is consistent with their temperature stability. Using presently available components, the drift characteristic would be substantially reduced.

Several developments within the last two years have shown continued improvement. By using a six point network and successive computer calculations it is shown that for the AN/PRC-70 tactical tranceiver the stability shown in Fig. 70-191-3 may be achieved. For temperature ranges restricted to 20° centigrade a stability of one part in 10^8 is possible.

There have been many papers on oven development including the use of Dewar flasks to reduce the power required to maintain control. One in particular is of interest in that it employs an idea proposed many years ago. Heat is supplied directly to the quartz plate by depositing a carbon film on the periphery including the sensor control. Fig. 70-157-7 shows characteristics obtained with this approach. On the average about .6 of a watt is required at ambients as low as -40°C. This is about one-tenth that required for the very efficient Dewar method. Care must be exercised to design against temperature transients since gradients within the quartz plate can produce substantial frequency changes.

Crystal Filters

The first symposium paper on crystal filters appeared in 1956. It served to show the characteristics that could be obtained from conventional half-lattice designs using crystal units then available. Fig. 56-339-4 shows the attenuation obtained using 8 and 4 pole half-lattice designs.

Wide bands with low distortion over the band can be produced using inductance to increase the separation of the poles and zeros. Also letting some of the critical frequencies fall outside the pass region improves the in-band resistance in accordance with image-parameter theory. Fig. 58-500-5 shows the attenuation of a five crystal lattice section shunted by inductance to widen the pass band to 150 kHz at a center frequency of 20 MHz. The unwanted responses show up above the band because all crystal units were in a single lattice structure.

In the next few years, a number of papers were presented on crystal filters including ladder types using equal inductance crystal units and types with doubly resonant high frequency crystal units where the resonance separation was controlled by the plate dimensions.

It wasn't until several years later, in 1966, that the real break-through came in crystal filter techniques. Researchers at the University of Tokyo, The Bell Laboratories and the General Electric Co. each independantly presented papers at the 20th Symposium on monolithic crystal filters. The Japanese approach is shown in Fig. 66-266-2. It shows a pair of trapped energy resonators with a finite separation together with an equivalent lattice structure for design purposes. Measured transmission characteristics of single and tandem pairs are shown in Fig. 66-266-4. The Bell approach is shown in Fig. 66-288-10 where multiple trapped energy resonators are coupled on a single quartz plate. Both the lattice and ladder equivalents are shown for design purposes. A summary of measured characteristics normalized to a single chart are shown in Fig. 66-288-12. The attenuation peaks shown for the dual resonator case are formed by capacitive coupling from input to output terminals. The loss characteristics for the three and six resonator cases are monotomic as shown by the dashed and dotted lines. The General Electric approach is shown in Fig. 66-309-6 where a pair of trapped energy resonators is shown with an equivalent circuit indicating coupling between the resonant circuits in general terms. The transmission characteristics of several filters of the dual resonator type for different electrode separation and hence band-width are shown in 66-309-10. The introduction of capacity from input to output terminals to provide attenuation peaks was also demonstrated.

The final link on the monolithic filter to modern network synthesis was achieved as a result of investigations to determine a practical equivalent circuit by W. P. Mason and W. L. Smith. With this approach it is possible to obtain any of the Tschebyscheff, Butterworth etc. responses. The design of an n-pole monolithic filter could then be obtained from the published tables of Dishall, Weinberg, Zverev and others.

The transmission loss characteristic of a six-pole filter is shown in Fig. 67-179-12 for both measured and calculated points. Note the close correlation for losses up to the 60 db level. Above this level the measured loss drops off due to scatter propagation. An X-ray topograph of a six pole filter is shown in Fig. 67-179-8 showing the zeros or critical input frequencies for a shorted output. These may be looked upon as the fundamental and anharmonic overtones of the entire electrode array. It is easy to distinguish the first, second, etc. to the sixth modes.

An alternative approach is to use capacitively coupled dual resonators. The response of a four section (8-pole) and 5 section (10 pole) filter is shown in Fig. 68-188-9. This method takes advantage of the cascading principle and eliminates the problem of scatter propagation and also unwanted modes.

Fig. 69-79-9 shows the computed effect of resonator Q on the in-band characteristics of an eight resonator monolithic filter. This is for a 10.7 MHz filter with a band-width of 3 kHz. Reasonably good response shown by the third curve from the bottom requires a Q of 200,000.

Eight pole monolithic filters, under production conditions for broadband multiplex show the characteristics given in Fig. 70-84-10. The lines with dots give the average and standard deviation for 24 filters plotted against the computed loss curve. The results for the in-band loss are shown in Fig. 70-84-11. This matches very closely the results indicated in a figure shown just previously by an entirely different group of authors.

Three other applications of coupled resonators have been presented. Fig. 68-206-6 shows three resonators adjusted to different frequencies and connected to a diode network to perform the function of a discriminator. This eliminates the necessity for an additional modulator to low frequencies to obtain a stable discriminator. Impedance transformation may be obtained as shown in Fig. 68-206-9 by the introduction of a capacitor with the associated retuning of the end resonator. Finally a stable low frequency source may be derived from a three resonator array as shown in Fig. 68-206-10. The difference frequency of the two end resonator controlled oscillators is detected by the diode connected to the middle resonator. Since all resonators are on the same quartz plate, temperature tracking should be much better than preselection of individual crystal units.

Future Developments

Now let us take a look at the future and see what we may anticipate. First, the business picture because this is easily disposed of. We asked a few of our friends what they thought about it and to one, we asked "What is your five year plan?" He replied that since it had been recently scrapped and substituted by a survival plan he was in no position to talk. Based on what has transpired businesswise over the past 25 years, the quality will go up, the accuracy will go up and the price will go down. And to summarize this picture, the price and delivery war will be fought on the local battleground with conventional weapons. It appears there will be little change in basic techniques; names, languages and acronyms will be different but fundamental concepts will remain the same.

Now, let us look at the technical possibilities. What are the trends? In communications systems, we have analog and digital systems, both of which are dependent upon an accurate frequency source. In the digital case we need a clock which in many cases must be regenerated and/or phased locked with the transmitting or a central clock. In analog systems, in addition to the frequency generator or clock, we need a precise filter to select one channel from another. With the present progress in integrated circuitry, frequencies can be generated very simply and synthesized to place them anywhere in the frequency spectrum that is needed. All that is required is one precise generator. Very probably there will be fewer crystal units produced for these systems but they will have higher accuracy.

For the higher frequencies there should be a material size reduction. Present 10 MHz crystal units can be produced with crystal plates 7.5 m m in diameter or less. At 30 MHz, they should be 2.5 m m or less. It is inconceivable that any of these plates would be put in present holders. They probably will be in the form of thin enclosures not over 30-50 mils thick and will have beam leads. It is believed that if we take a look at the progress that has been made in integrated circuit technology and apply some of those techniques to the present crystal design and developments, we would be far ahead. For example, with smaller size plates, batch processing becomes feasible and except for final adjustment, 100 vibrators can be processed from a 1 inch square plate. There should be some standardization of generator frequencies and of crystal filters in order to make them off-the-shelf items for applications to digital and analogue, wire and radio transmission systems. In the future, there is going to be more and more use of mobile and marine radio equipment.

With developments in active solid state devices moving in the direction of performing functions, it is inconceivable that elements for frequency control and selection would not follow this same trend. Some of this has already started but presently available VXCO's and filters resemble little the advancements in integrated circuitry. What is envisioned is a completely integrated series of functions such as: Frequency generators, synthesizers, translators and modulators; narrow band amplifiers for SSB, AM and FM systems, as well as discriminators for FM systems. These functional devices should be of the size and performance to be compatible with other functions to form a basic equipment used in a system. So far as the other new concepts are concerned, some of them may still be buried in the symposia proceedings in a manner similar to that of Mortley's article disclosing the trapped energy concept published in Wireless World in 1951. The apple just didn't happen to drop on the right head at that time. The monolithic filter as we know it today could have been available more than 20 years ago. Either we were not ready for the concept or simply did not recognize it at the time. So far as other items in the future are concerned, we leave those to your imagination.

What we have just said, of course, is only an opinion and if we were to tell you what really would happen in the next 25 years, you wouldn't believe it!

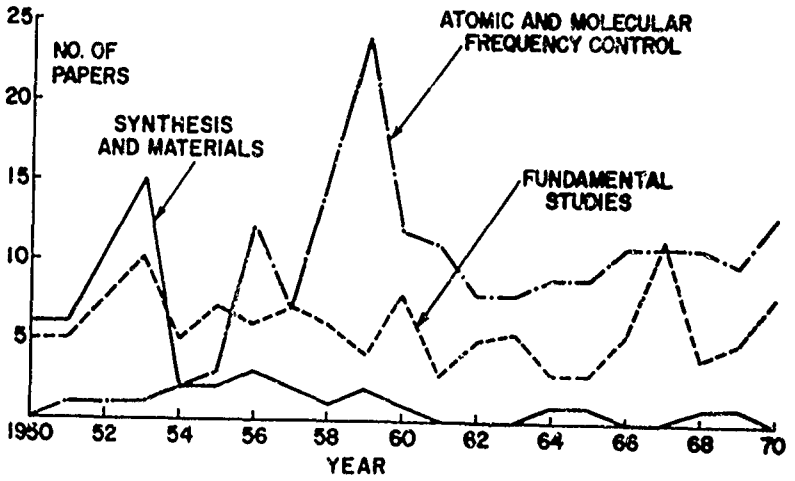


Fig. 1 - Number of papers given in fields of synthesis, fundametnal studies of srystals and atomic & modular frequency control.

CATEGORIES	NO. OF PAPERS REVIEWED
SYNTHESIS AND MATERIALS	8
FUNDAMENTAL STUDIES	22
MEASUREMENT AND TEST	13
CRYSTAL UNIT DEVELOPMENT AND ENGINEERING	26
TEMPERATURE CONTROL AND COMPENSATION	8
CRYSTAL FILTERS	15

Fig. 2 - Number of papers in various categories selected for review.

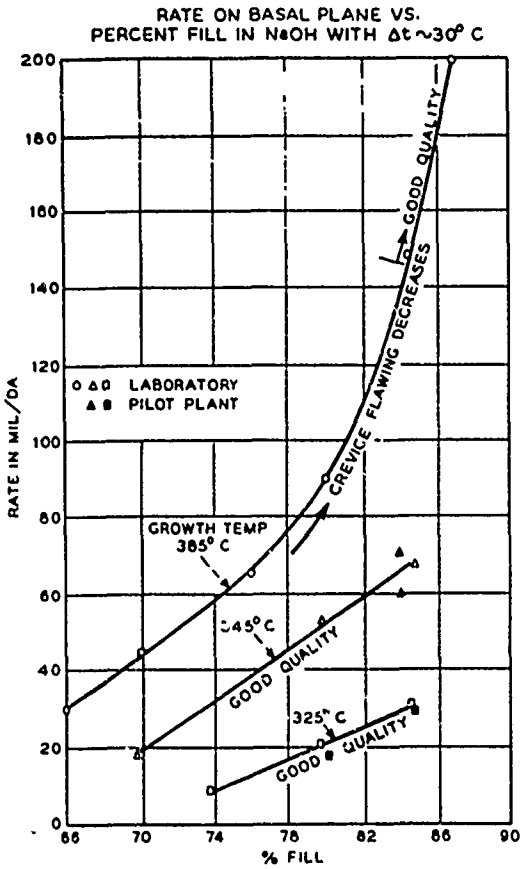


Fig. 3 - Growth rate of synthetic wuartz.

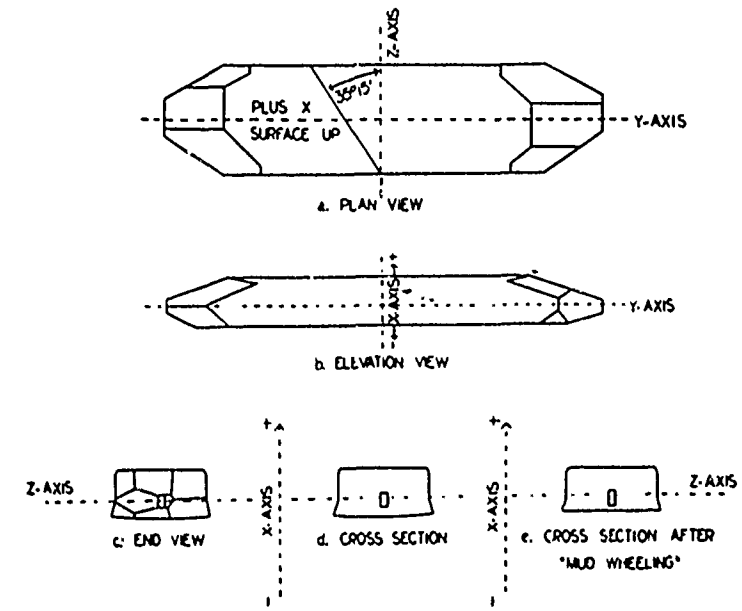


Fig. 4 - Synthetic quartz-bar growth.

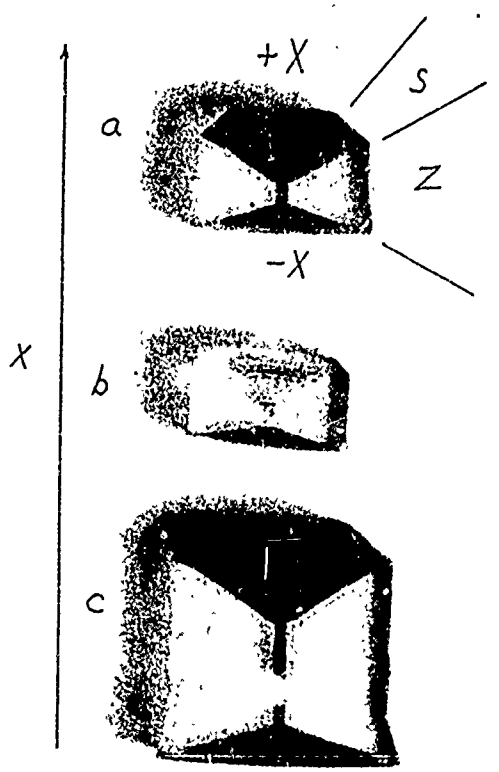


Fig. 5 - Irradiated sections of synthetic quartz.

MATERIALS	SYMBOL	CLASS		NO. PIEZOELECTRIC CONSTANTS
LITHIUM GALLIUM OXIDE	Li_1GaO_2	ORTHORHOMBIC	C_{2v}(2mm)	5
LITHIUM NIOBATE	Li_1NbO_3	TRIGONAL	C_{3v}(3m)	8
LITHIUM TANTALATE	Li_1TaO_3	TRIGONAL	C_{3v}(3m)	8
LITHIUM ALUMINUM OXIDE	YLi_1AlO_2	TETRAGONAL	D_4(422)	2
CALCIUM PYRONIOBATE	$Ca_2Nb_2O_7$	MONOCLINIC	C_{6v}(6mm)	5

Fig. 6 PIEZOELECTRIC MATERIALS UNDER STUDY

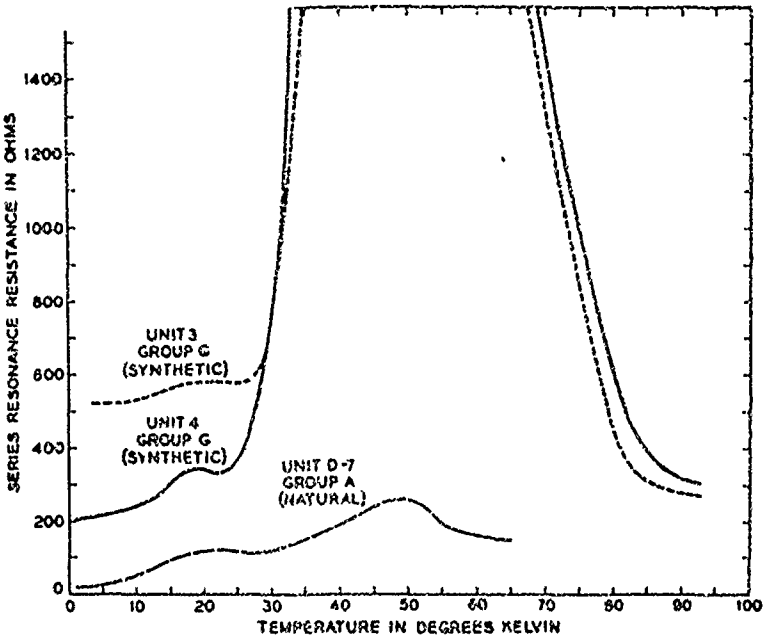


Fig. 7 - Loss peaks in natural and synthetic quartz.

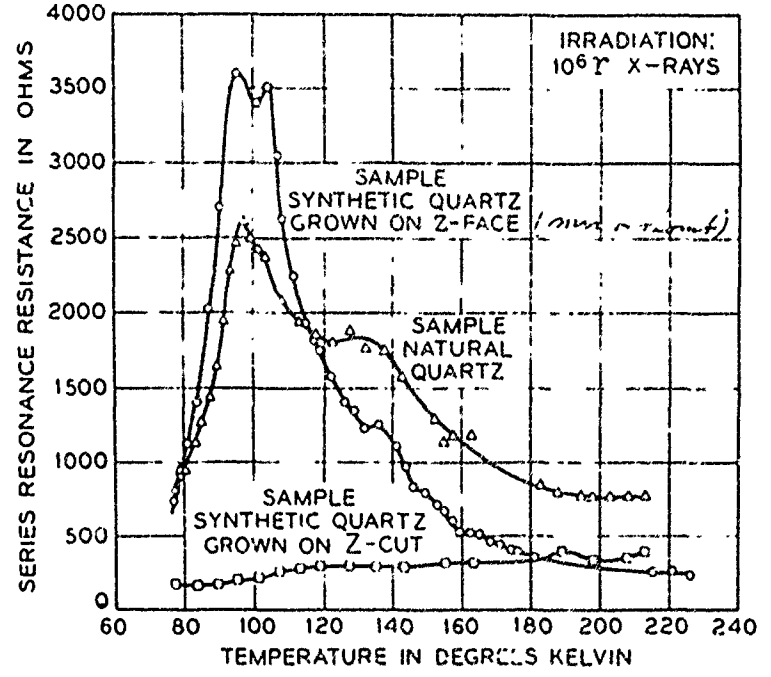


Fig. 8 - Loss peaks in irradiated quartz.

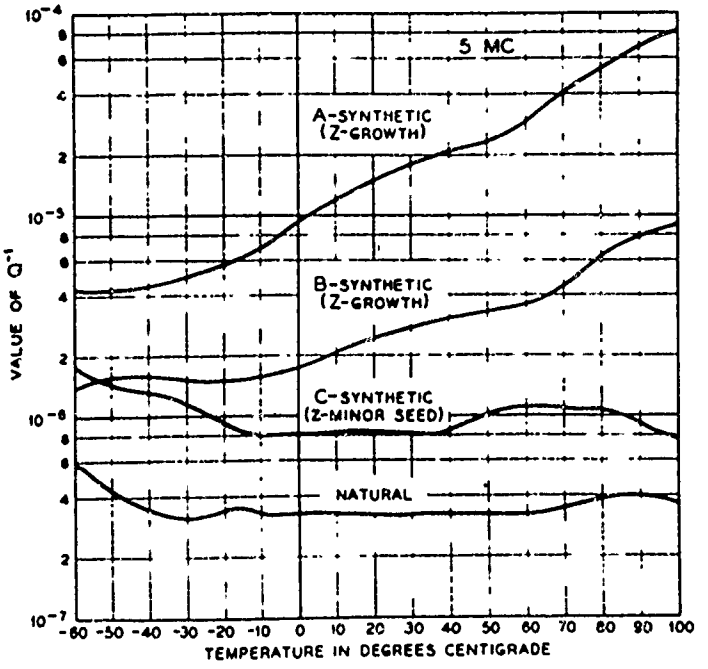


Fig. 9 - Acoustic absorption at 5 MHz between -60° and 100°C in some representative samples of synthetic and natural quartz.

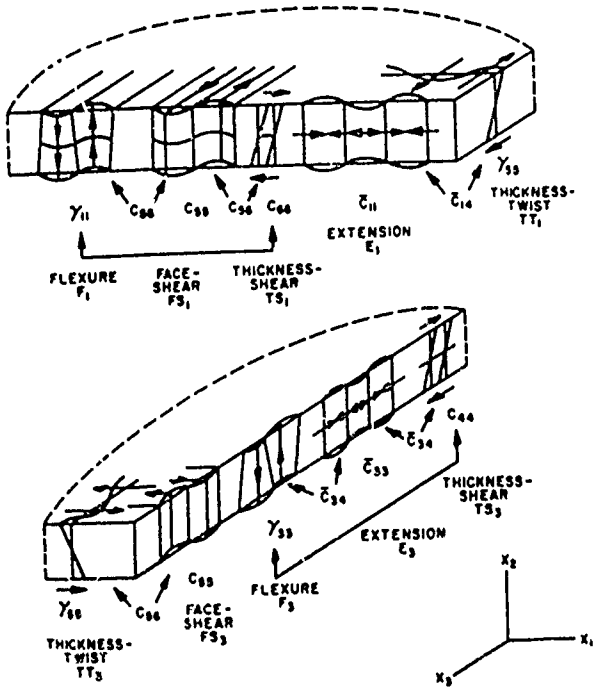


Fig. 10 - Mode shapes in quartz plates.

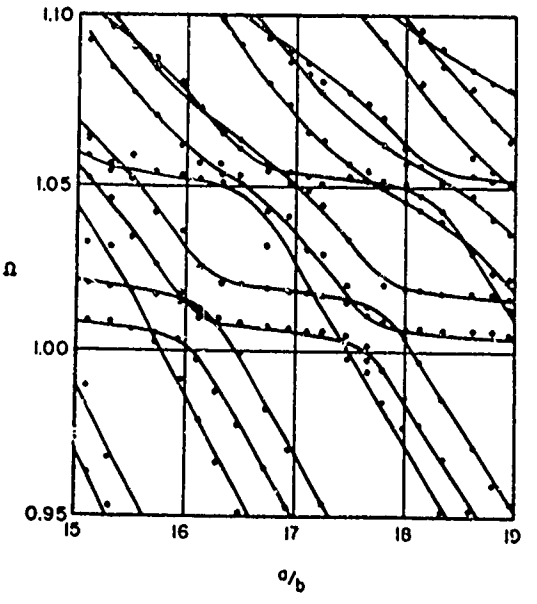


Fig. 11 - Computed and measured resonances for rectangular AT-cut quartz plates, length c/thickness b=20.

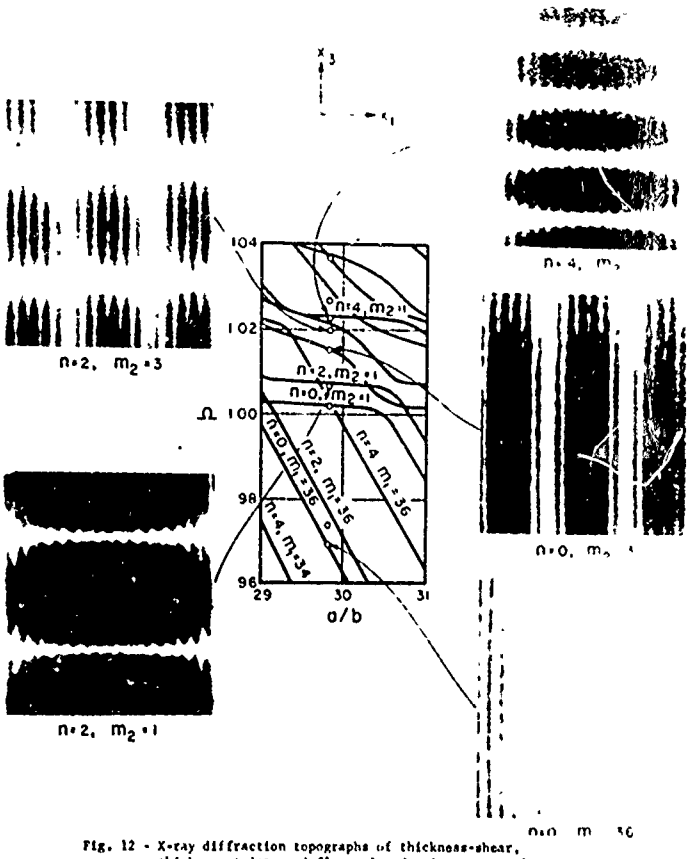


Fig. 12 - X-ray diffraction topographs of thickness-shear, thickness-twist, and flexural modes in rectangular AT-cut quartz plate $\frac{a}{b}$ = 29.8.

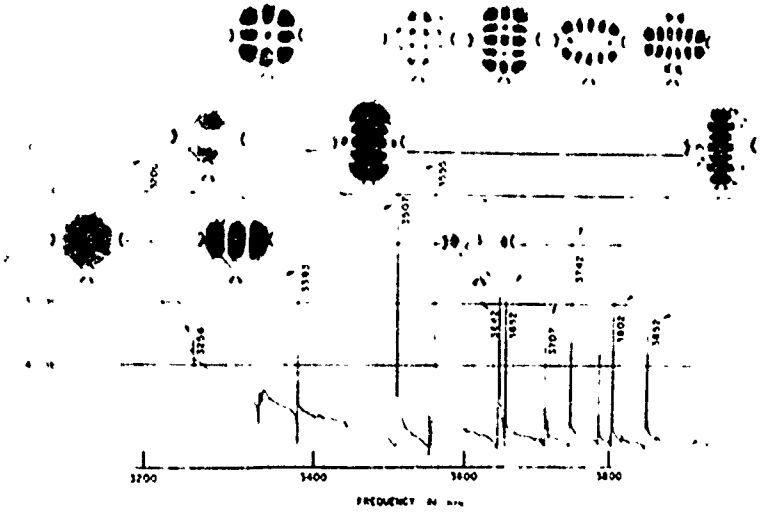


Fig. 13 - Topographs of modes in AT-cut quartz plates.

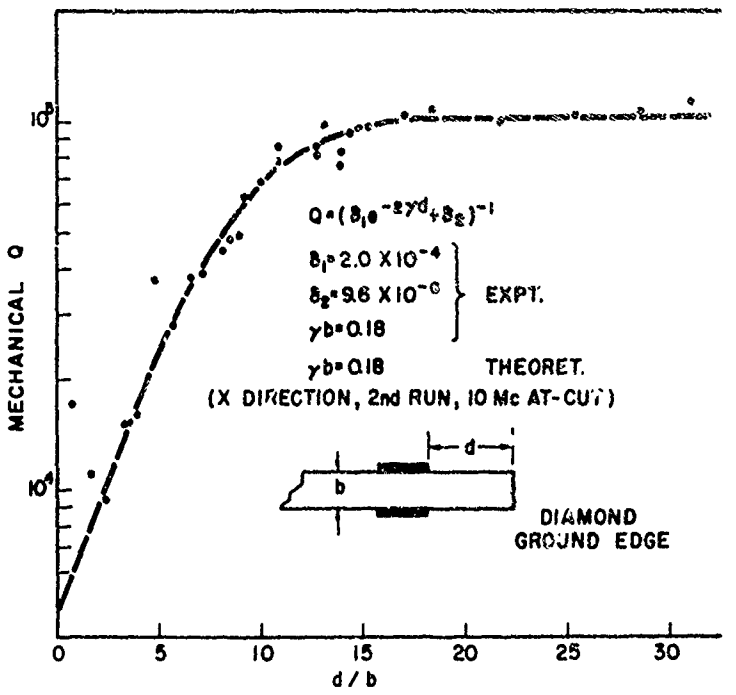


Fig. 14 - Mechanical Q as a function of distance to diamond ground edge of wafer.

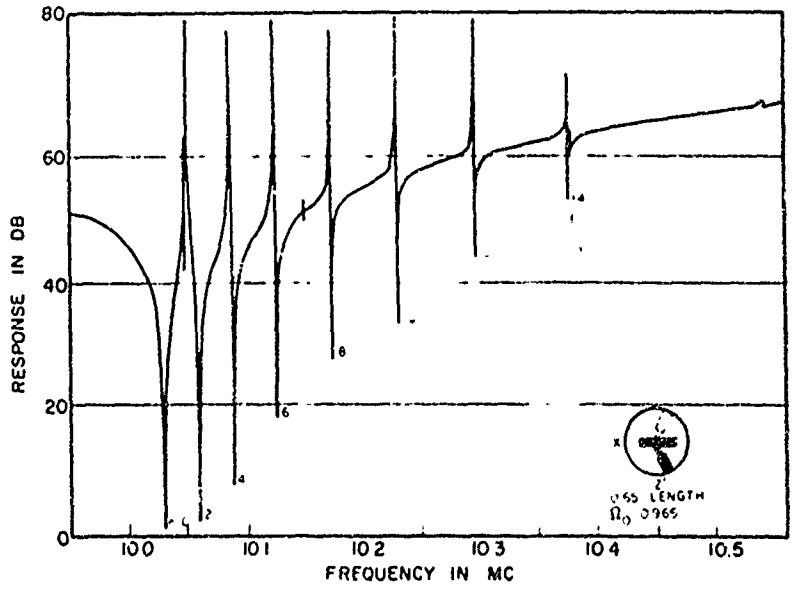


Fig. 15 - Resonator response with rectangular electrode in X direction, electrode length = 100b.

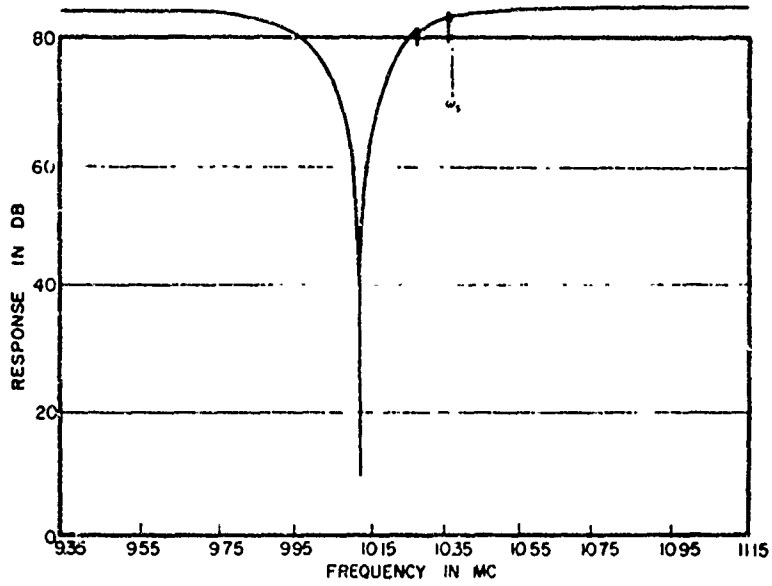


Fig. 16 - Resonator response with rectangular electrode, length 10b.

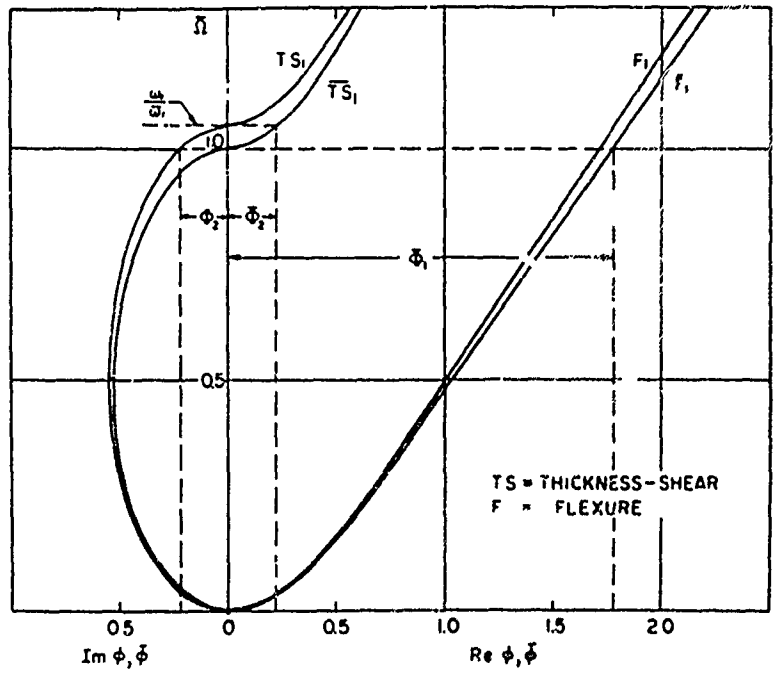


Fig. 17 - Dispersion curves for thickness shear and flexure in a quartz plate.

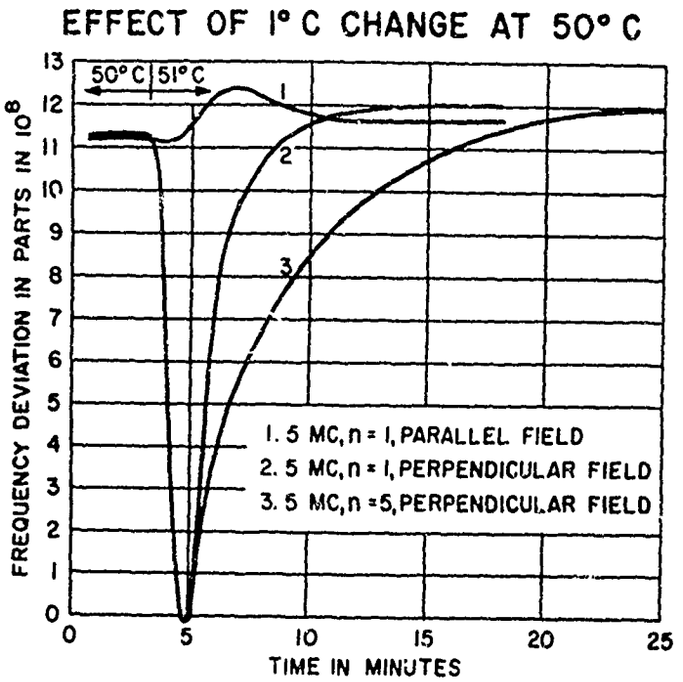


Fig. 18 - Frequency change of precision crystal resonators due to temperature gradient.

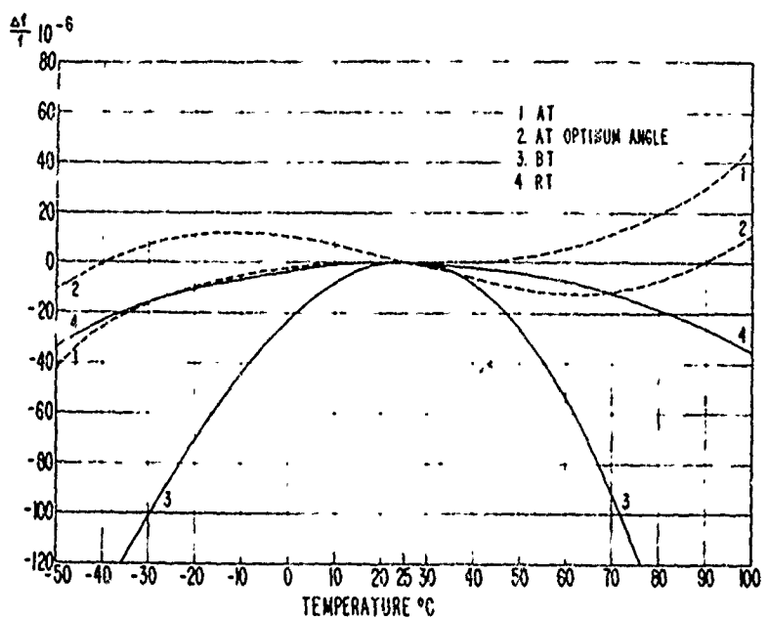


Fig. 19 - Frequency-temperature curves for AT, BT and RT-cut quartz resonators.

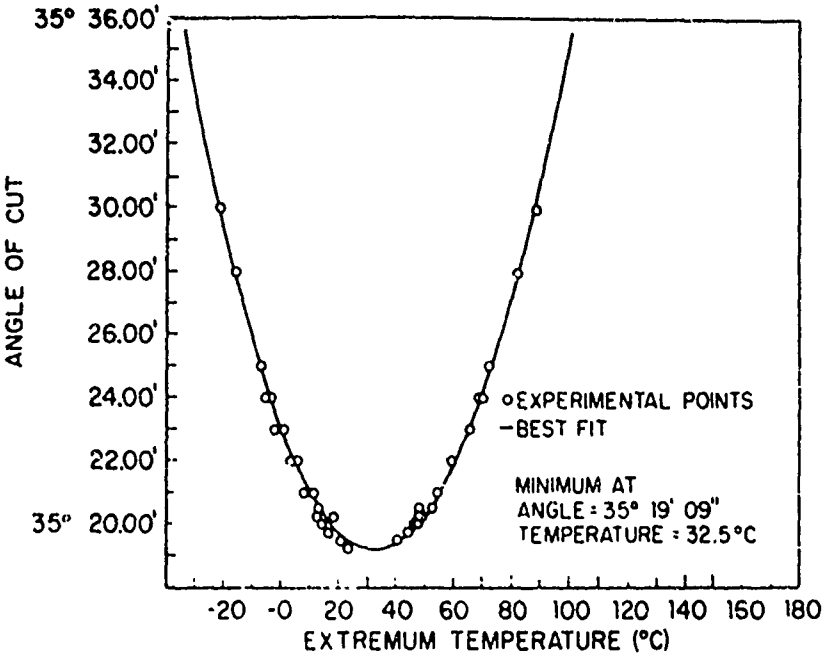


Fig. 20 - Locus of turn-over temperature for AT-cuts.

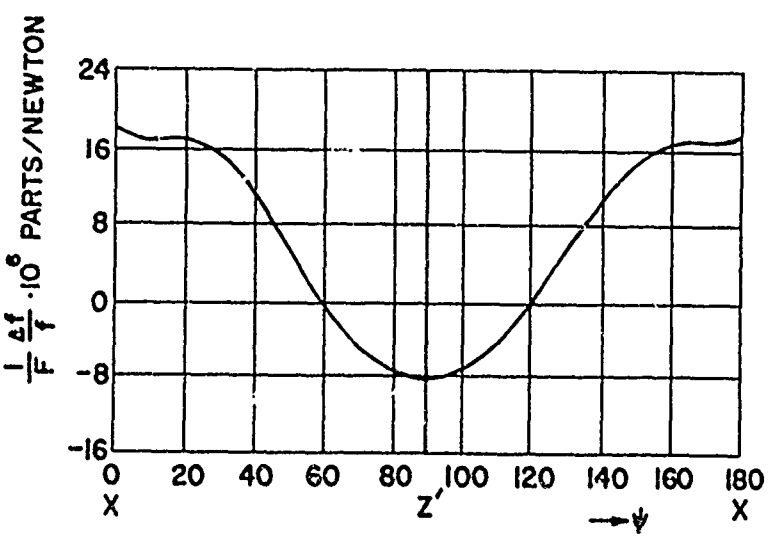


Fig. 21 - Relative frequency change per unit of applied radial force for AT-cut resonators.

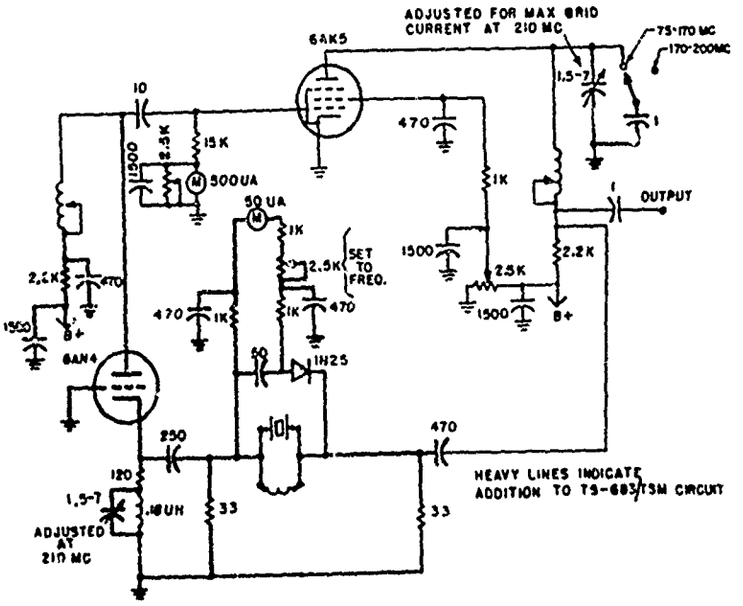


Fig. 22 - Simplified circuit of AN/TSM-15 Test set.

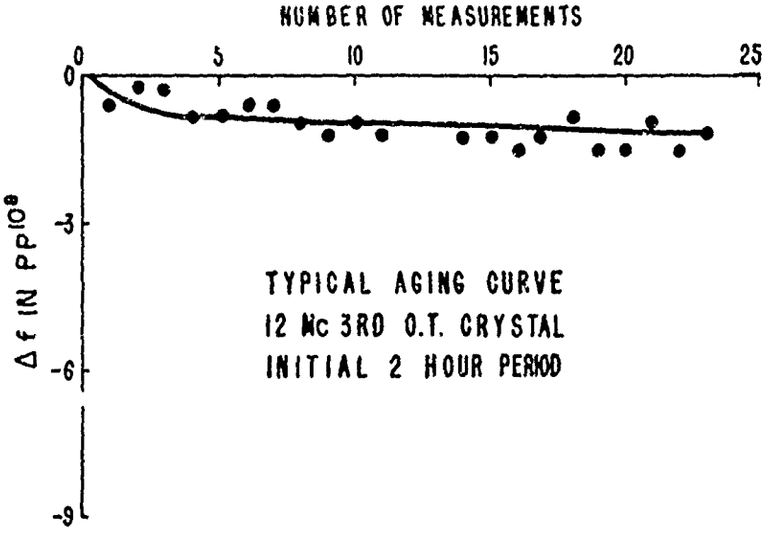


Fig. 23 - Typical accuracy curve, 12 MHz 3rd overtone crystal unit.

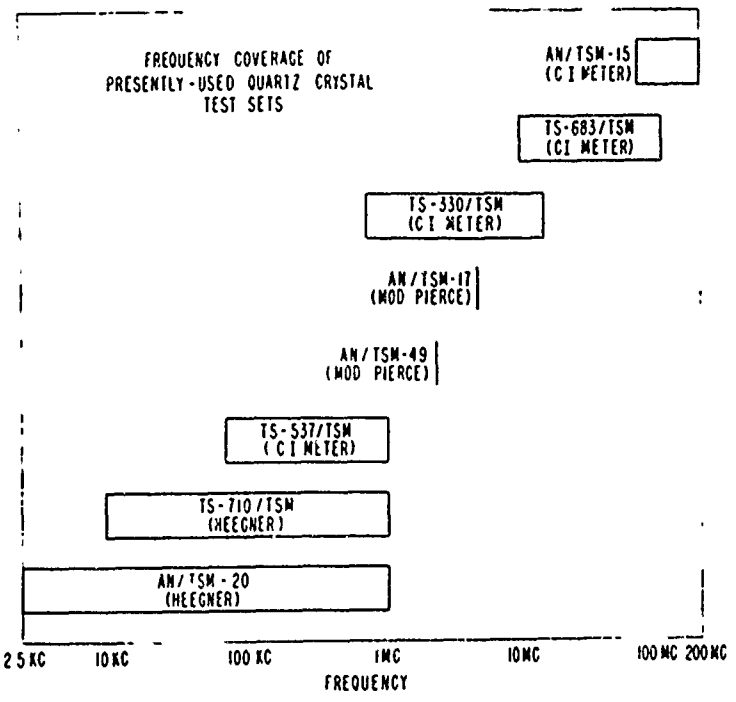


Fig. 24 - CI meters available or under development in 1964.

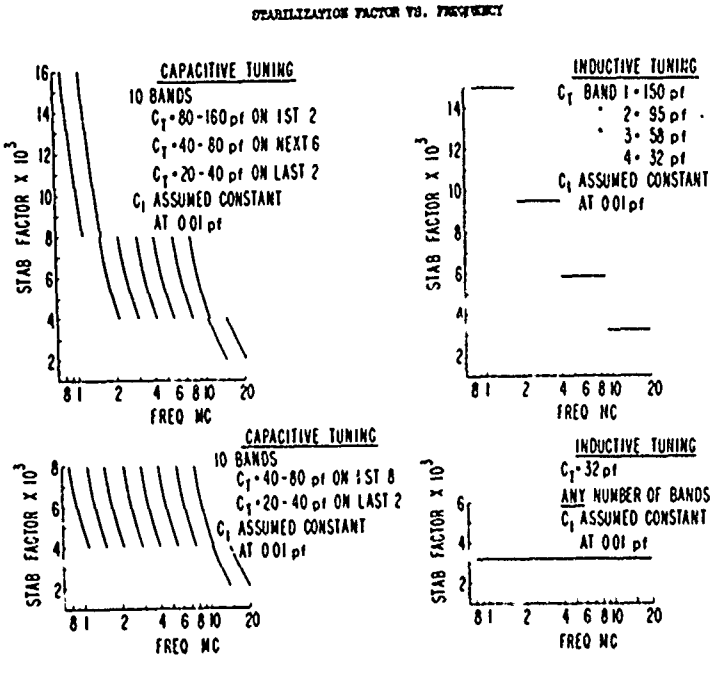


Fig. 25 - Stabilization factor vs. frequency.

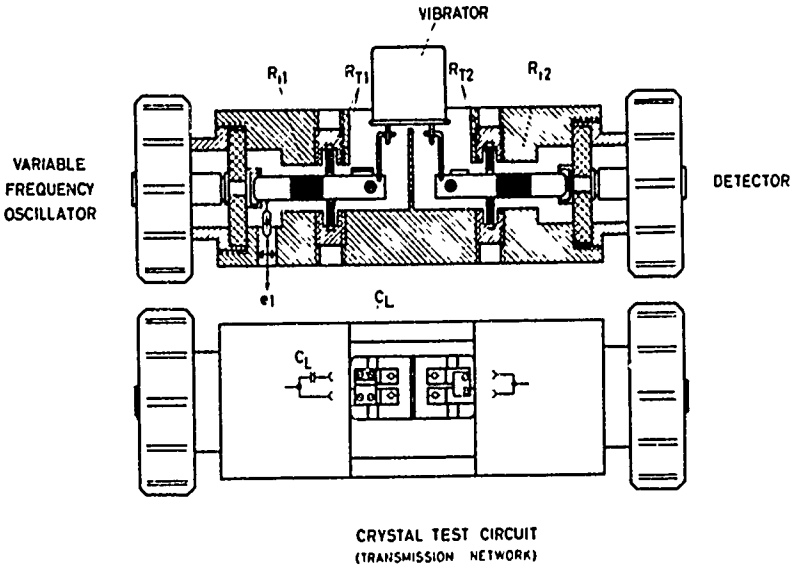


Fig. 26 - Crystal unit transmission test network.

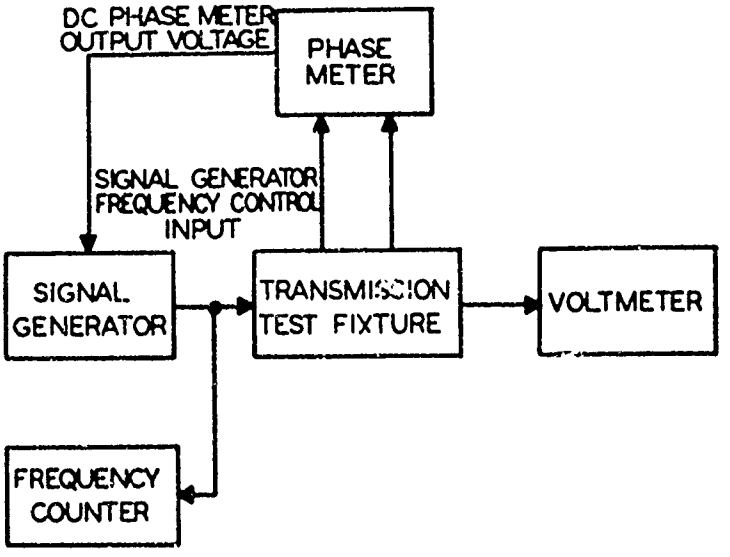


Fig. 27 - Block diagram of self tuned phase detection measuring system.

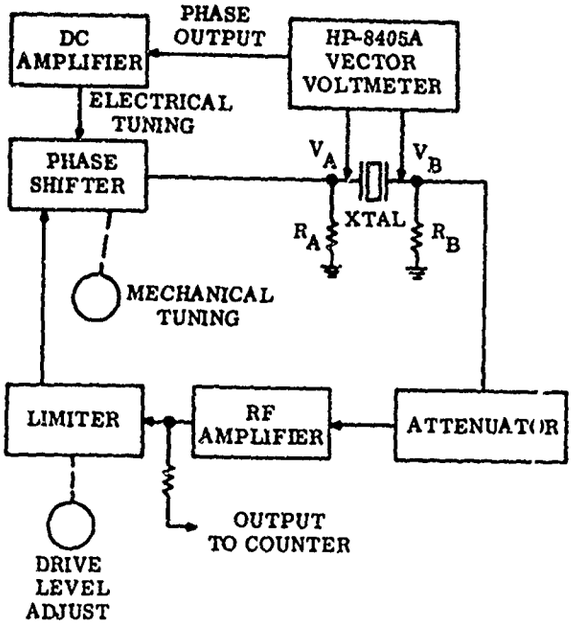


Fig. 28 - Block diagram of measuring system using vector voltmeter.

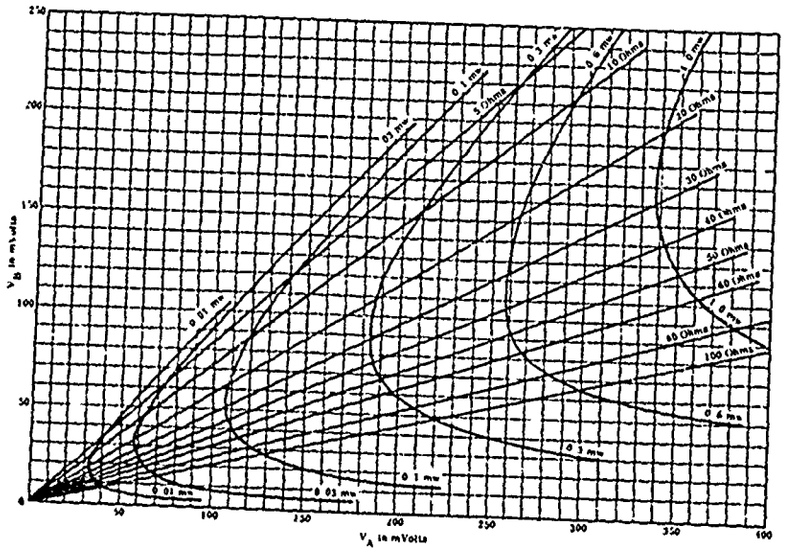
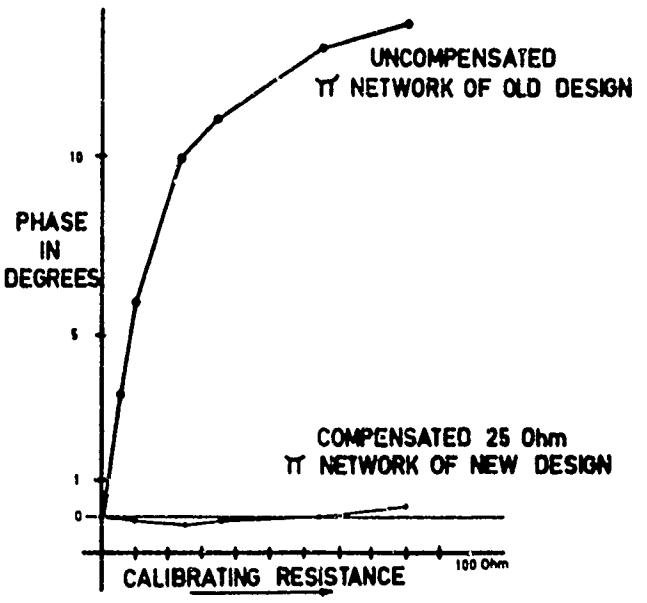
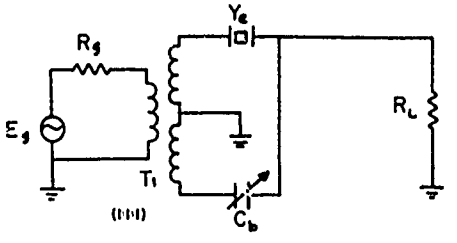


Fig. 29 - Resistance and drive level graph.

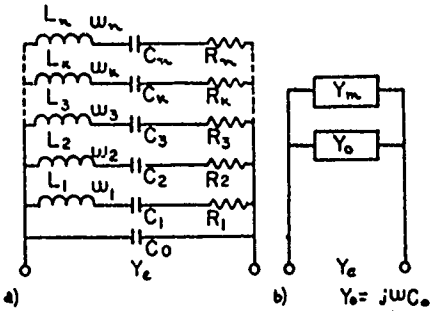


INFLUENCE OF DESIGN OF π NETWORK ON PHASE ERRORS.

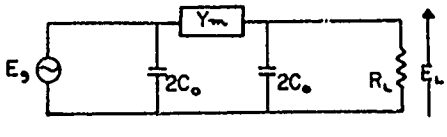
Fig. 30 - Influence of design of π networks in phase errors.



SCHEMATIC OF TEST CIRCUIT



EQUIVALENT CIRCUIT OF CRYSTAL VIBRATOR



EQUIVALENT CIRCUIT OF TEST CIRCUIT
(C=Co;T, IDEAL)

Fig. 31 - Test circuit for measurement of unwanted modes.

Fig. 32 - Pressure Rise Detection Methods

Test Method	Type Unit	Sensitivity
Frequency & ER	All	Dependent upon resonator cut
Tesla Coil	Glass	5×10^{-2} to 5 Torr
Hot Wire Pirani	All	$\geq 10^{-2}$ Torr
Bead Thermistor	All	3×10^{-4} to 10 Torr
Ionization Current	Metal	3×10^{-1} to 10 Torr

Fig. 33 - Actual Leak Test Methods

Test Method	Trace Substance	Min. Detectable Leak
Bubble Test	Air	10^{-5} std cc/sec
Bubble Test	Helium	10^{-6} std cc/sec
Pressurized Die and Vacuum	Die Penetrant	See Note 1
Mixed gas back-filled	Helium	10^{-10} std cc/sec
Accumulation	Helium	10^{-11} std cc/sec
Accumulation	Argon	10^{-13} std cc/sec
Radioactive tracer gas	Radene (Krypton 85) (and Nitrogen)	10^{-13} std cc/sec (10^{-10} to 10-9 std cc/sec go-no-go

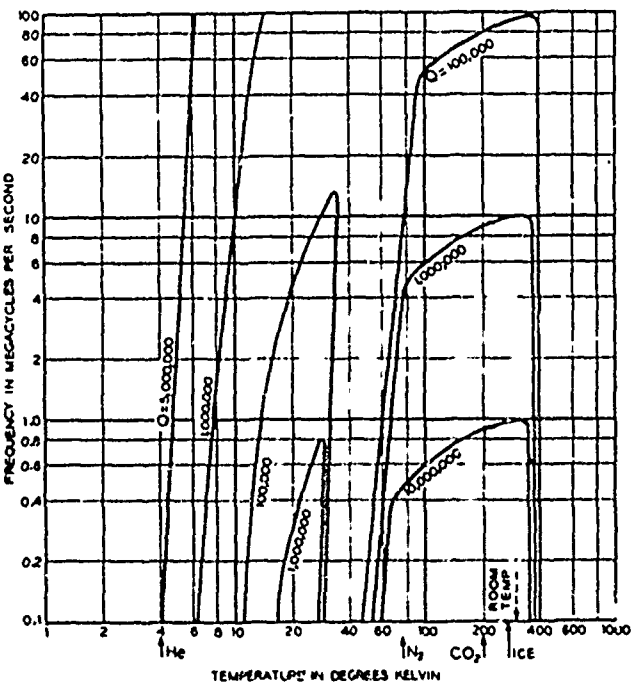


Fig. 34 - Variation in Q with temperature and frequency due to intrinsic loss in quartz.

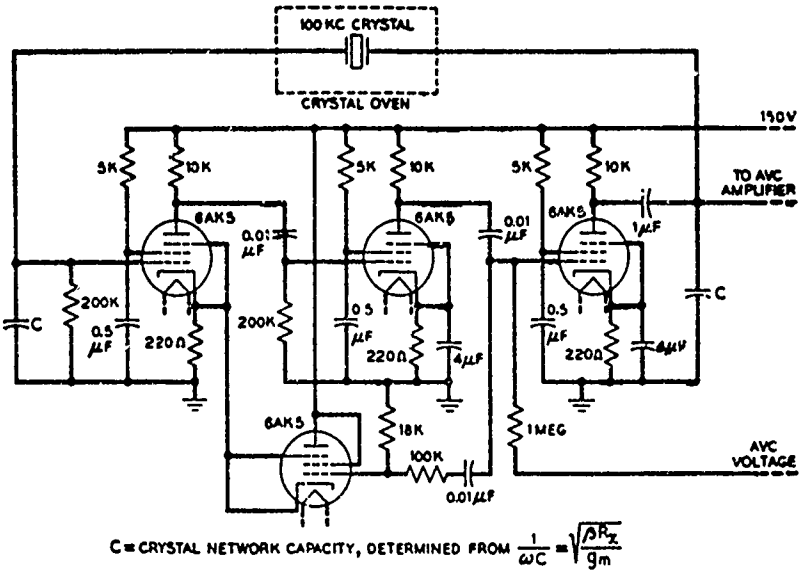


Fig. 35 - Multi-stage oscillator.

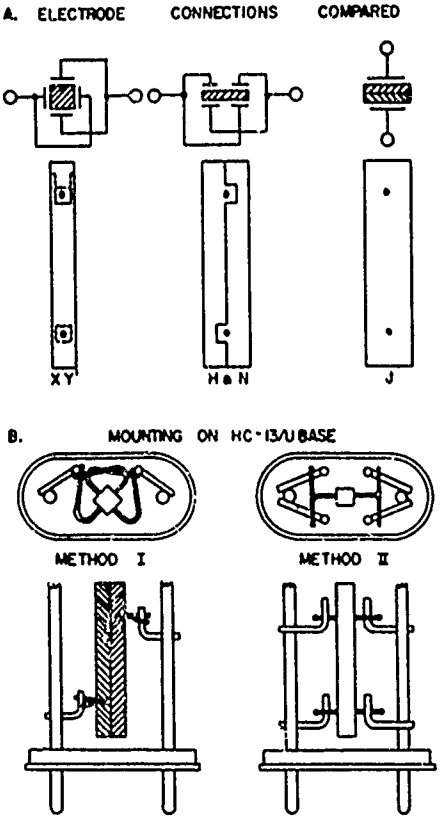


Fig. 36 - Flexure mode crystal unit designs.

	HIGH PRECISION	HIGH PRECISION	STANDARD CRYSTAL UNIT
Temp. Coef. (in 10^{-7}/°C)	1	2	10
Q (in 10^4)	100 - 200	10 - 50	5 to 20
Stability/hr (in 10^{-10})	1 - 10	10 - 100	1000
Stability/wk (in 10^{-8})	1	20	100
Drive Level (in uw)	10 to 100	500	2000 to 10,000
Price	x	$\frac{X}{2}$ to $\frac{X}{4}$	$\frac{X}{50}$ to $\frac{X}{100}$

Fig. 37 - Parameters and cost ratios of crystal units.

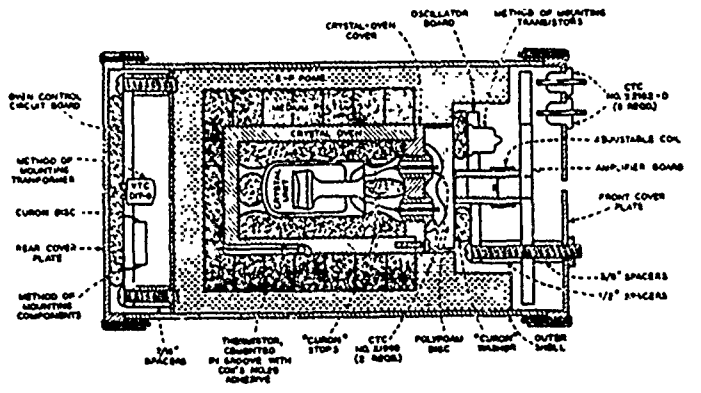


Fig. 38 - Oscillator for severe environments.

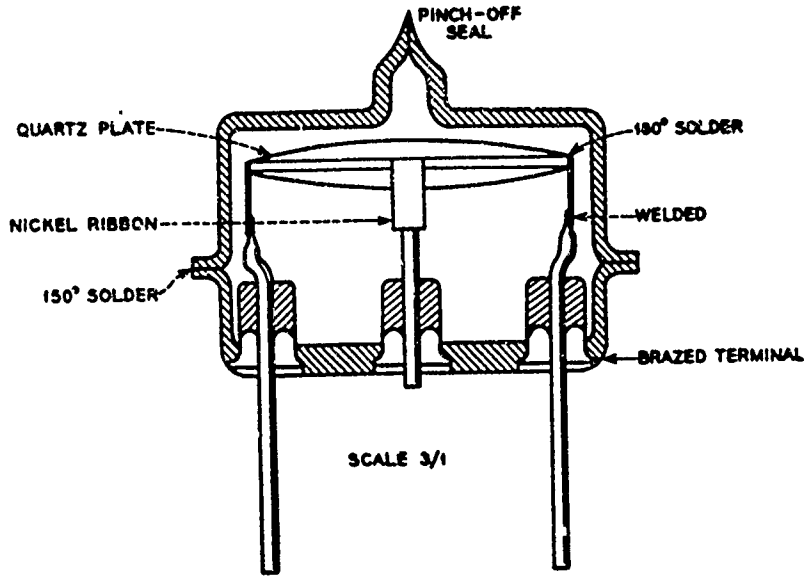


Fig. 39 - Crystal unit design for high acceleration.

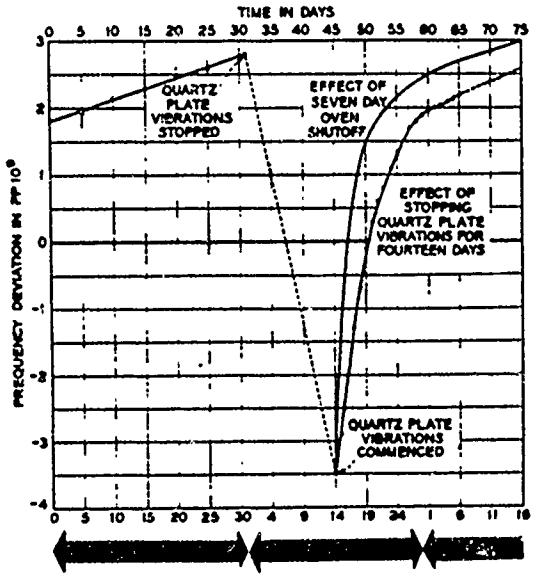


Fig. 40 - Effects of oven and osc. shutdown.

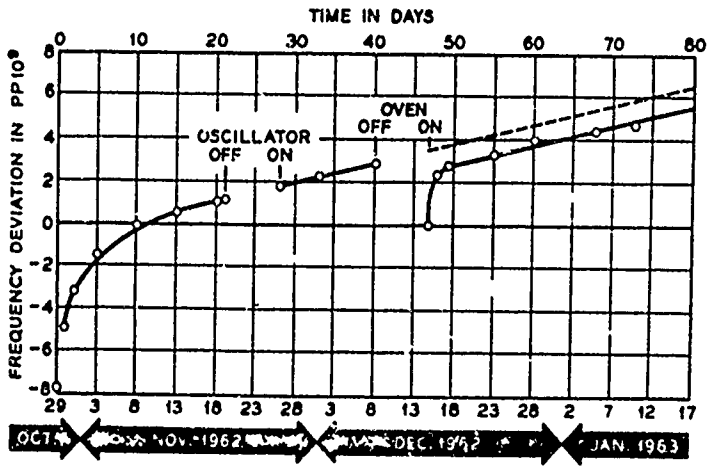


Fig. 41 - Effect of oven and osc. shutdown.

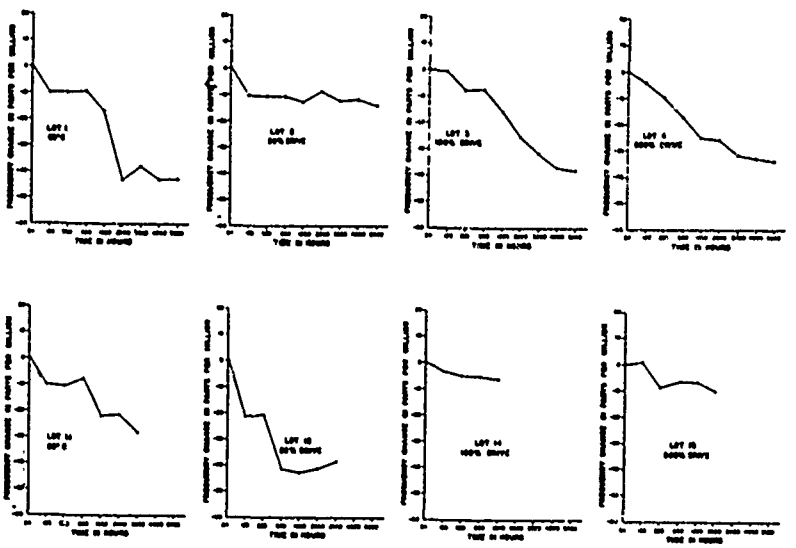


Fig. 42 - Aging of metal enclosed crystal units.

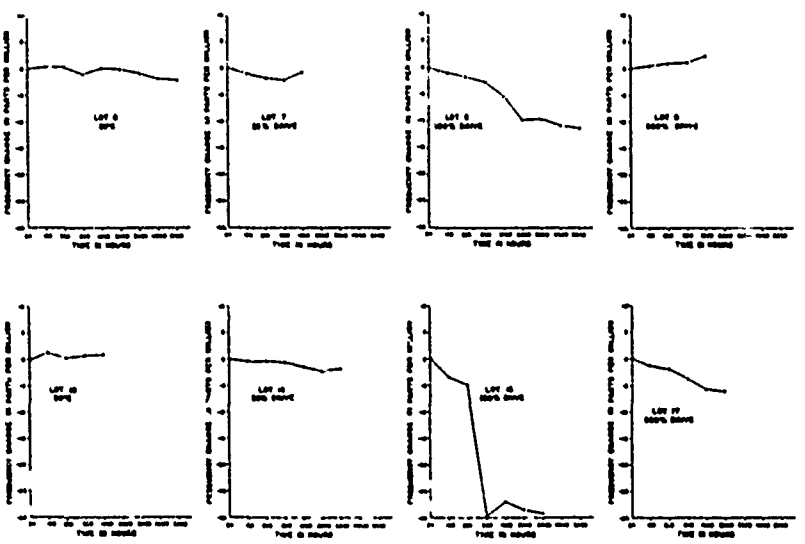


Fig . 43 - Aging of glass enclosed crystal units.

Fig. 44

COMPARATIVE FAILURE RATES* OF REPRESENTATIVE ELECTRONIC COMPONENTS

COMPONENT	FAILURE RATE (% / 1000 HRS)
CRYSTAL, QUARTZ	0.14
DIODES	0.298
ELECTROLYTIC CAPACITORS	0.248
MICA CAPACITORS	0.046
RESISTORS	0.036 TO 0.140
TRANSISTORS	0.103
TUBES	
RECEIVING	0.942
TRANSMITTING	4.29
MAGNETRONS	54.4

* FAILURE IN THE FIELD - IN SHIPBOARD AND LAND-BASED EQUIPMENT

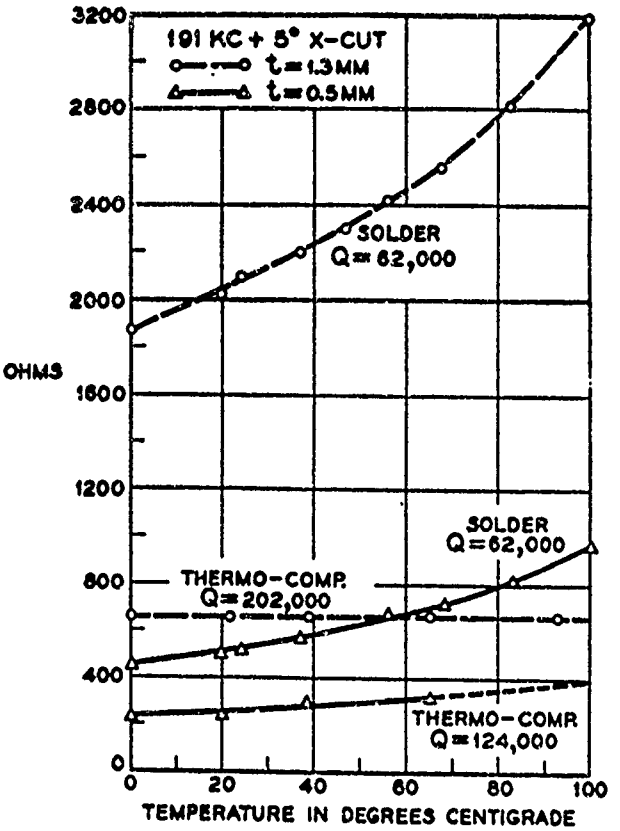
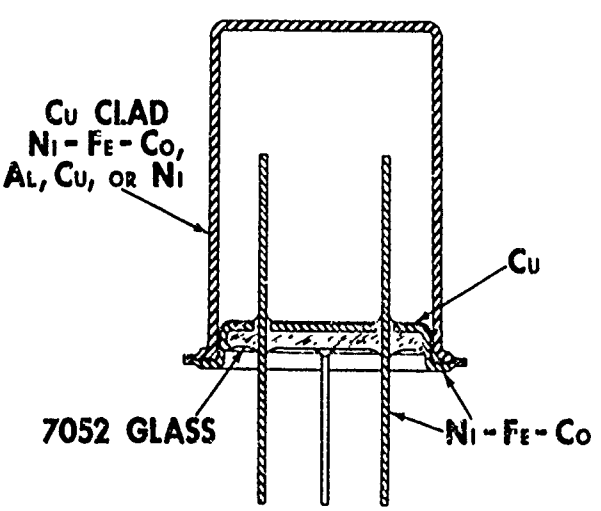


Fig. 45 - Resistance change in solder and thermocompression bonded crystal units.

Fig. 46

COLD WELDED QUARTZ CRYSTAL ENCLOSURE



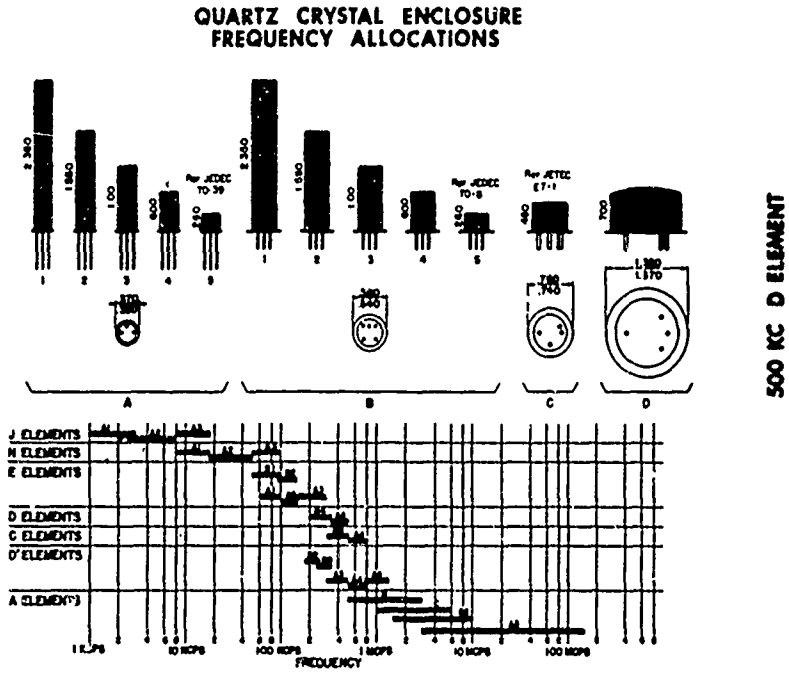


Fig. 47 - Crystal unit standard enclosures.

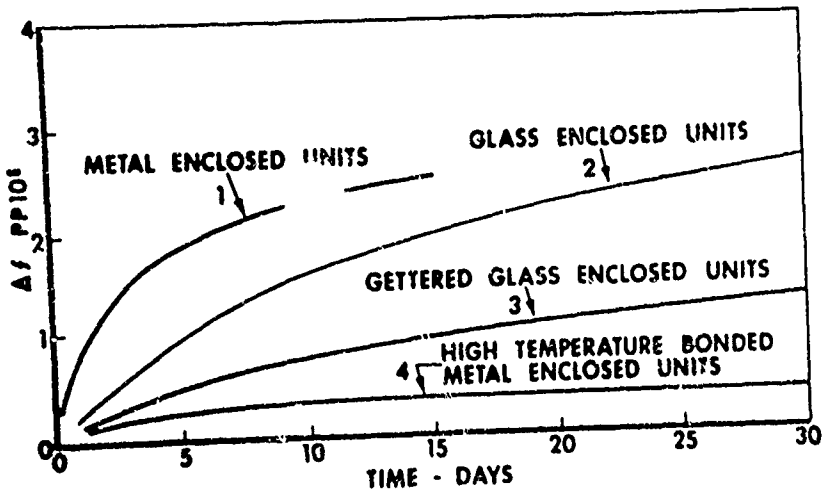


Fig. 48 - Aging of metal and glass enclosed 5 MHz crystal units.

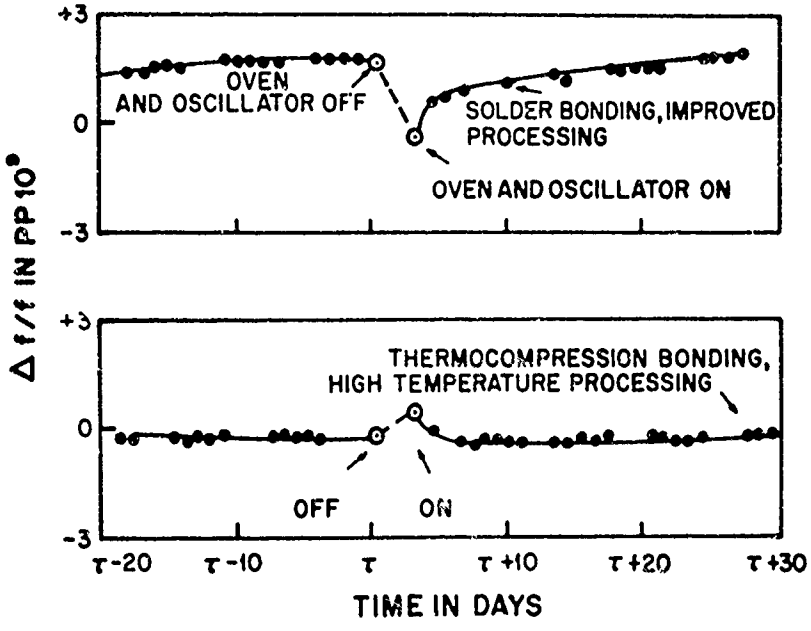


Fig. 49 - Frequency stabilization of precision 5 MHz crystal units after oven shutdown.

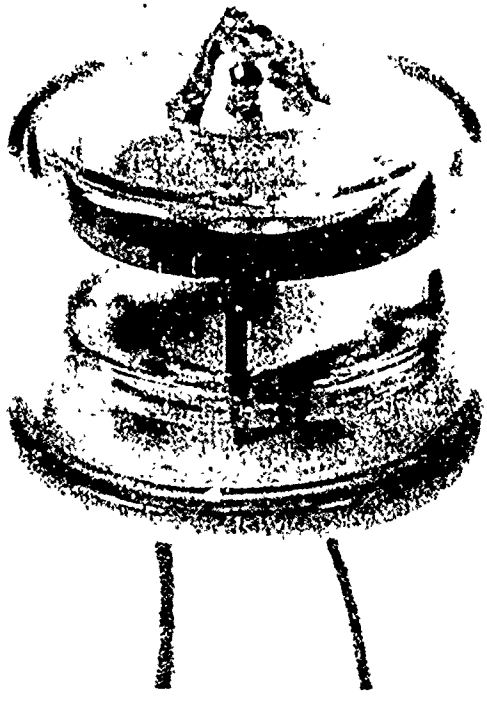


Fig. 50 - 5 MHz glass enclosed precision crystal unit.

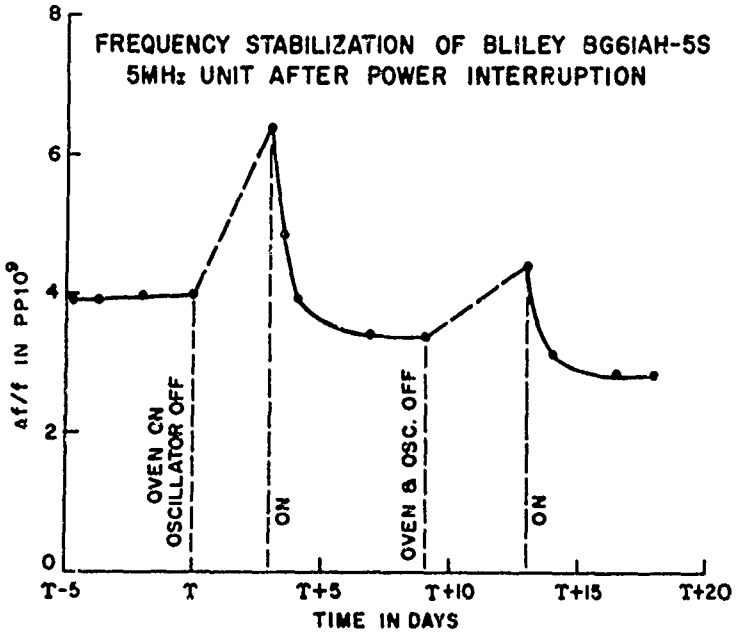


Fig. 51 - Effects of oven and osc. shutdown.

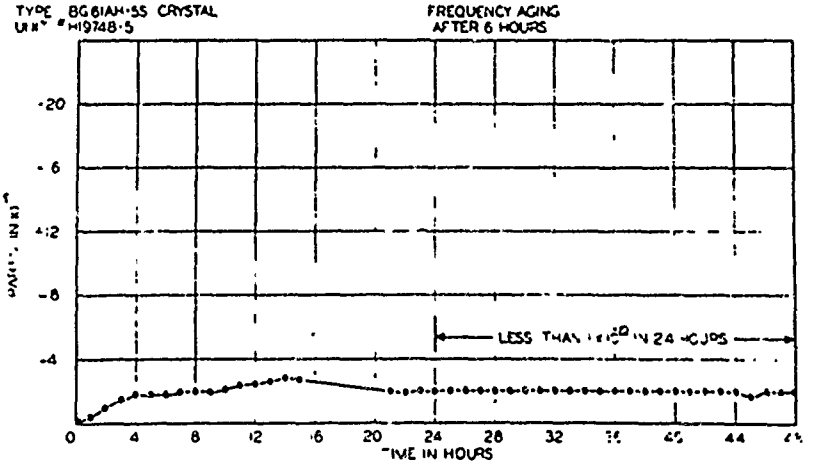


Fig. 52 - Aging characteristic of precision 5 MHz unit.

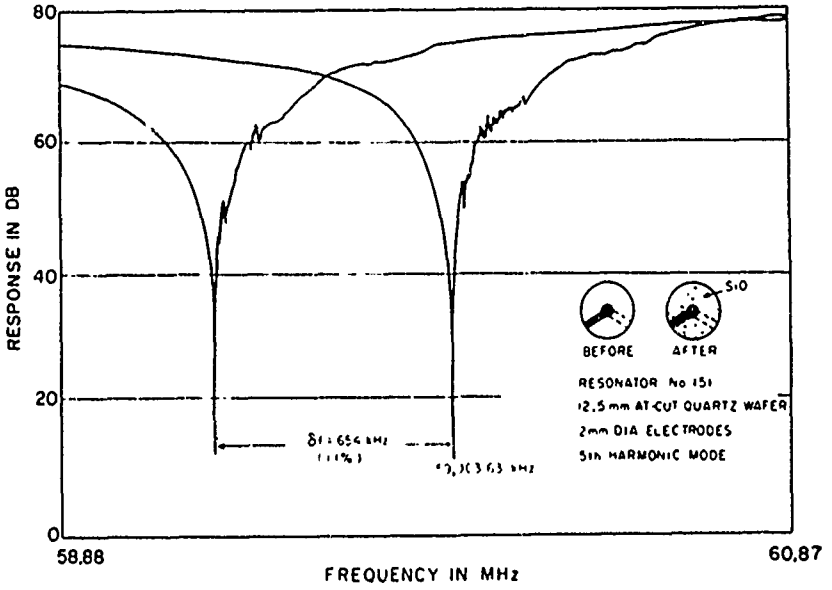
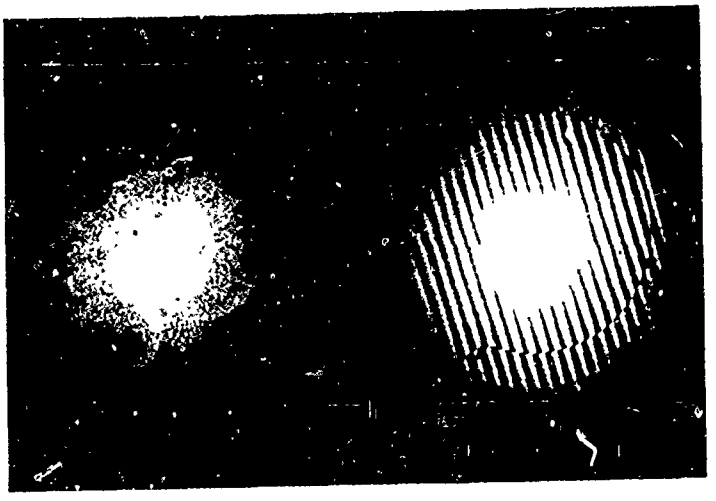
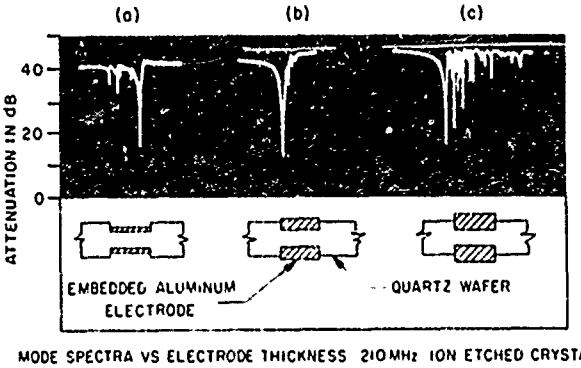


Fig. 53 - Effects of SiO dielectric tuning



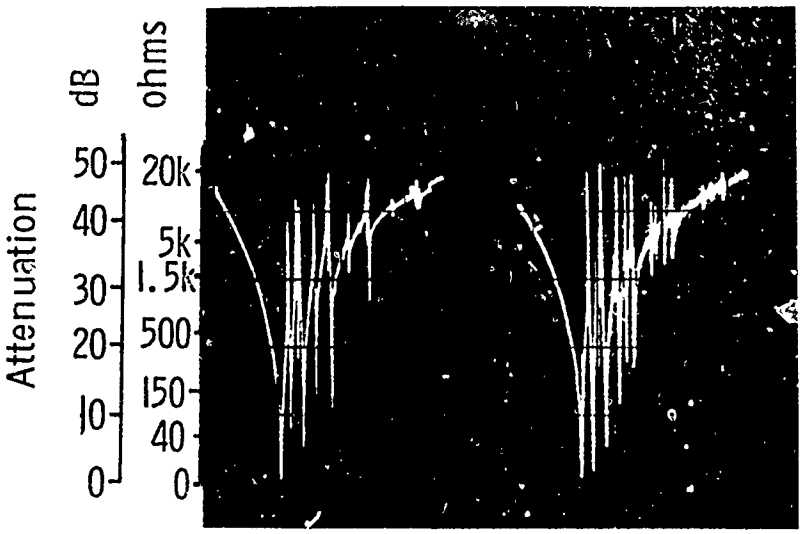
KEYHOLE PATTERN ERODED INTO A QUARTZ PLATE
BY ION ETCH TECHNIQUES

Fig. 54 - Use of ion etch techniques and resultant response characteristics.



MODE SPECTRA VS ELECTRODE THICKNESS 210MHz ION ETCHED CRYSTAL

Fig. 55 - Use of ion etch techniques and resultant response characteristics.



Typical 61 MHz Oscillator Crystal Mode Spectra

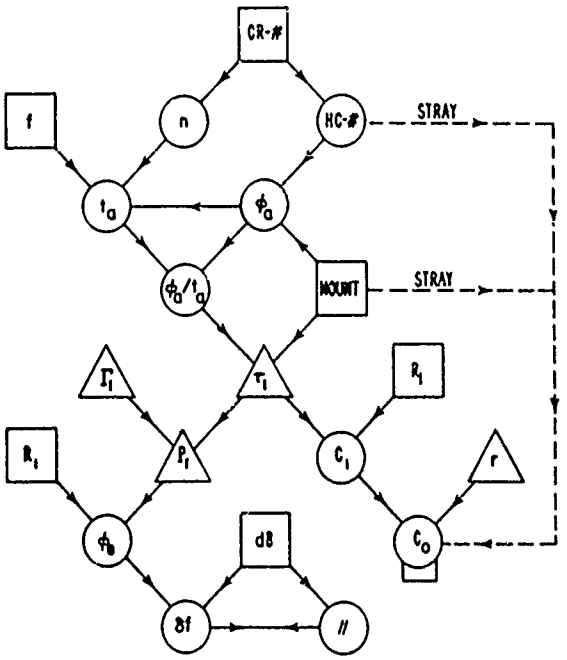


Fig. 57 CRYSTAL RESONATOR DESIGN ALGORITHM

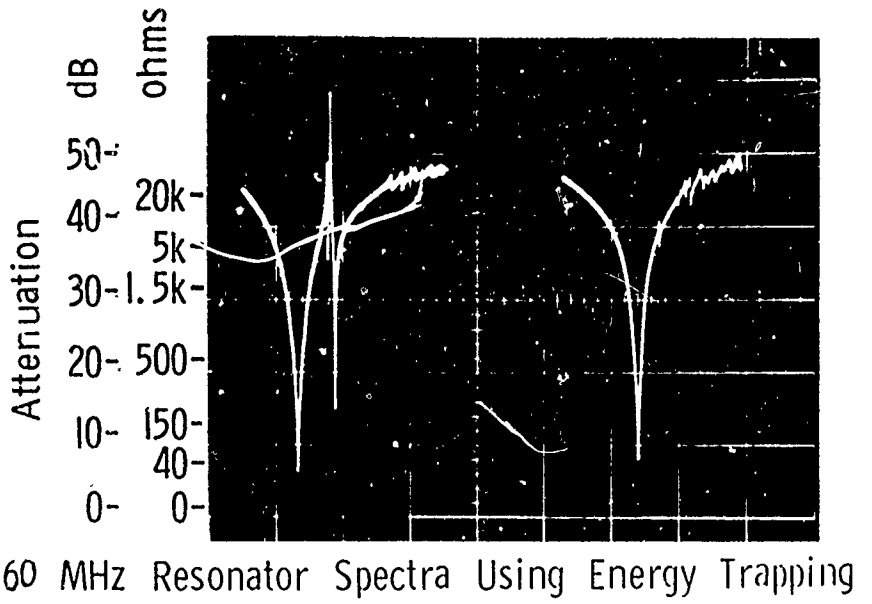


Fig. 58 - 60 MHz response spectra using design to eliminate trapped modes.

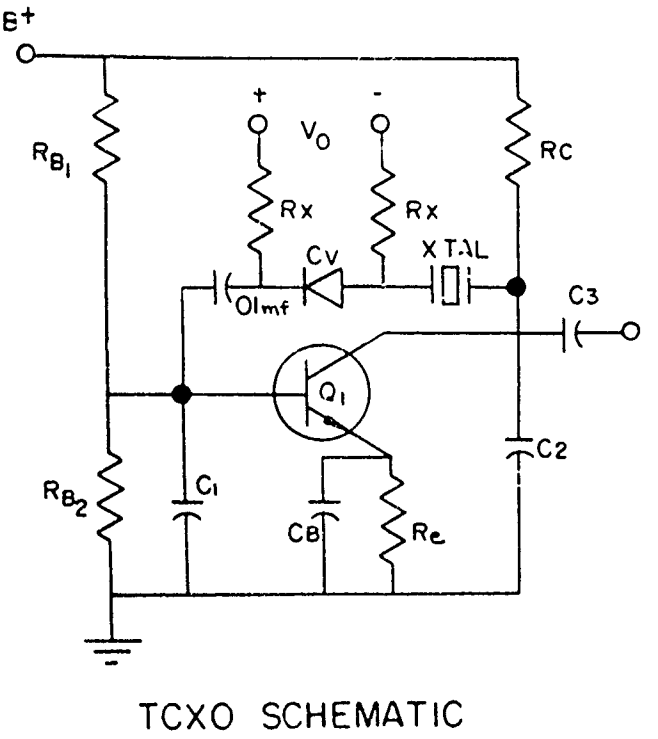
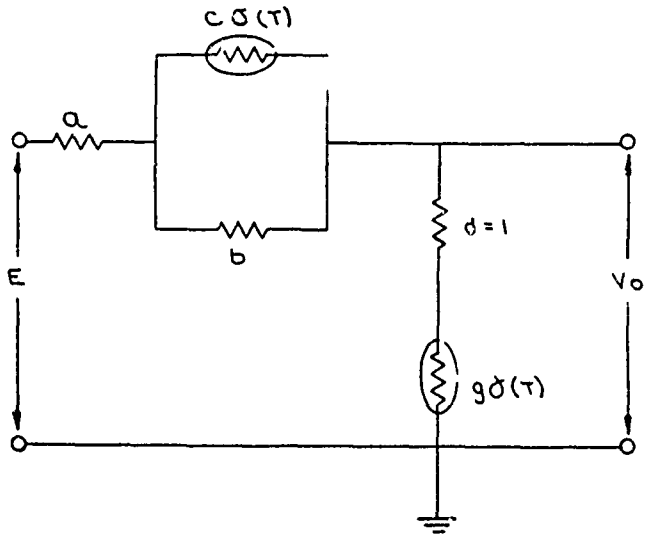


Fig. 59 - Compensated oscillator.



a, b, & c ARE FIXED RESISTORS. cσ(T) & gδ(T) ARE THERMISTORS WHERE c & g ARE Ro VALUES & σ(T) & δ(T) ARE TEMPERATURE DEPENDENT MULTIPLIERS.

Fig. 60 - Compensation network.

Fig. 61 CRITICAL AREAS OF TCXO DESIGN

1. Crystal Parameters
   - A. C_o/C_1 ratio
   - B. Δf/f between LTP and UTP
   - C. Turning point temperatures
   - D. Frequency tolerance at specified load capacitance
   - E. Aging characteristics
2. Varicap
   - A. Tolerance of C_x at 4 V
   - B. Leakage current
3. Compensation Network
   - A. Thermistor R_o and β tolerances
   - B. Resistor tolerance and T. C.
   - C. Error of calculated β to required β
   - D. Aging of thermistors and resistors
4. Oscillator circuitry
   - A. Gain stability with temperature and time
   - B. Temperature coefficient of components
   - C. Overall Q of circuit

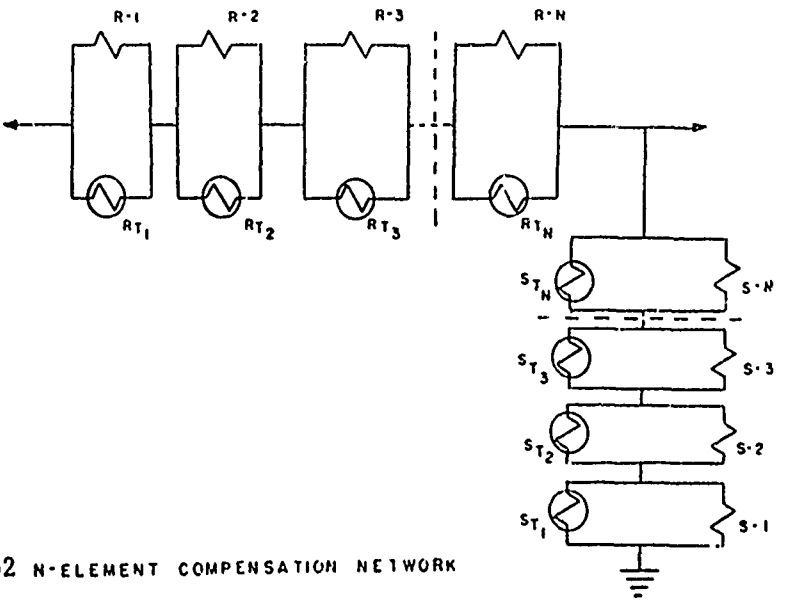


Fig. 62 N-ELEMENT COMPENSATION NETWORK

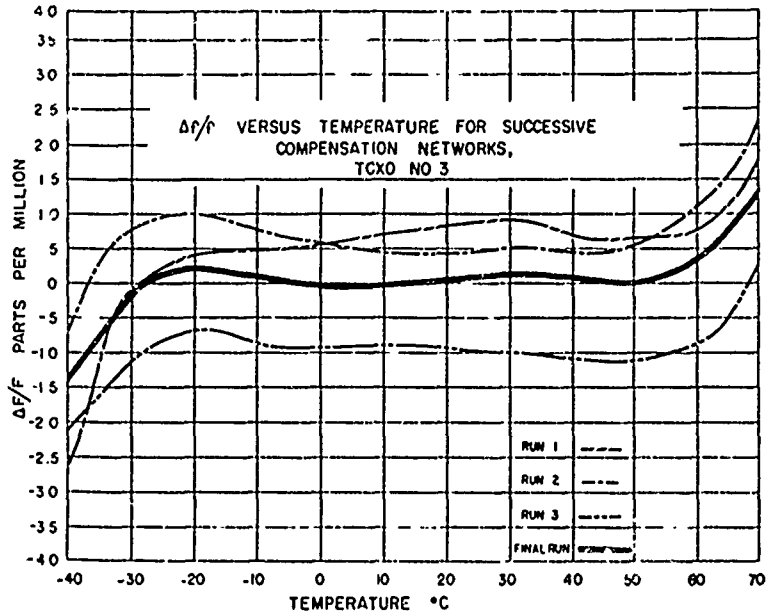


Fig. 63 - Results of successive compensation design.

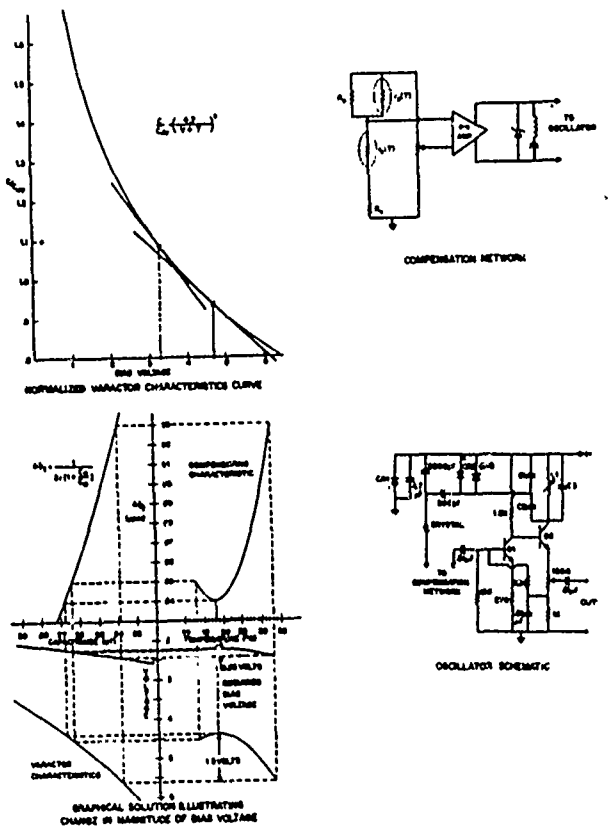


Fig. 64 - Compensation design factors.

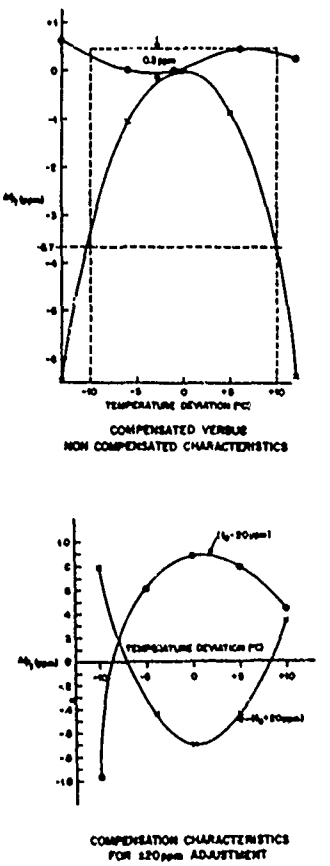


Fig. 65 Characteristics of compensated CT-units.

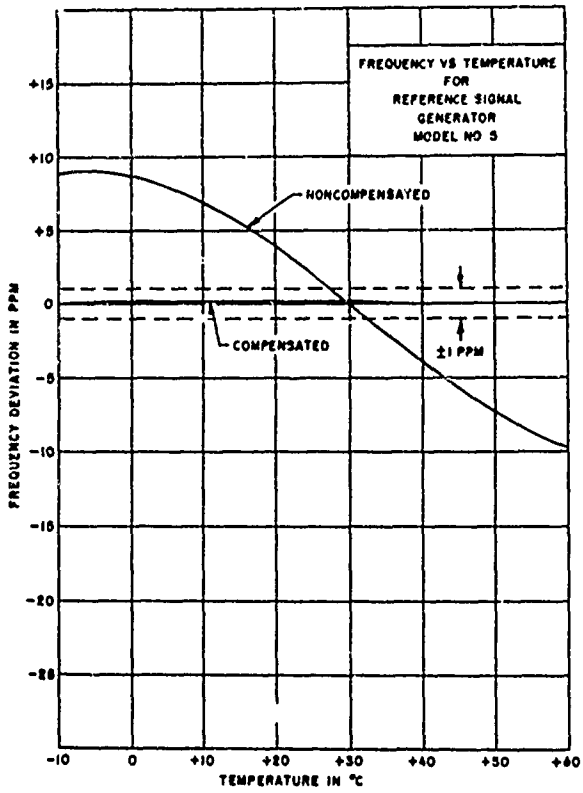
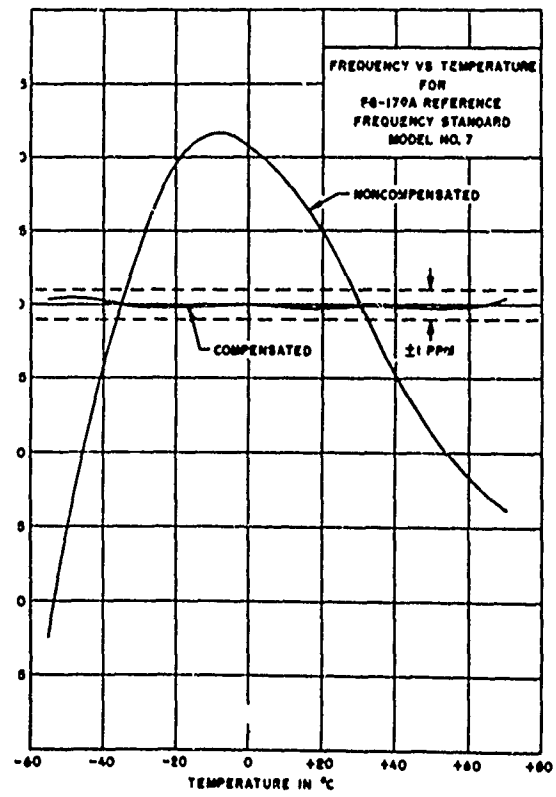


Fig. 66 - Frequency-temperature characteristics of the reference signal generator for AN/URC-32 transceiver.



Fig, 67 - Frequency-temperature characteristics for SG-179(XA-3) reference frequency standard.

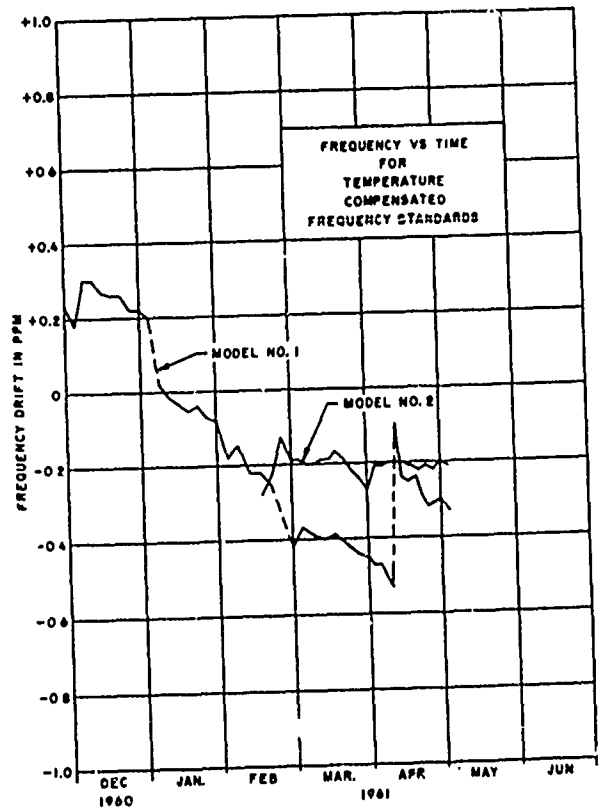


Fig. 68 - Aging for two temperature compensation frequency standards.

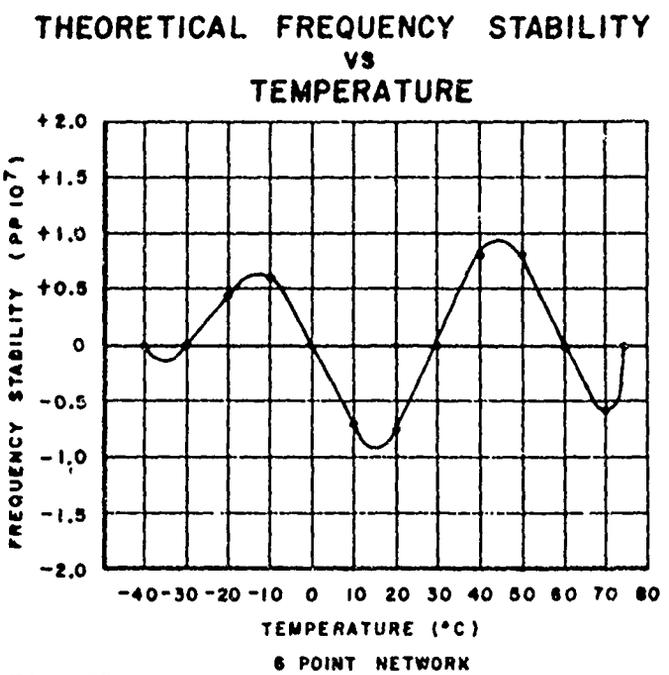


Fig. 69 - Frequency-temperature characteristic (computed) for a 6-point network.

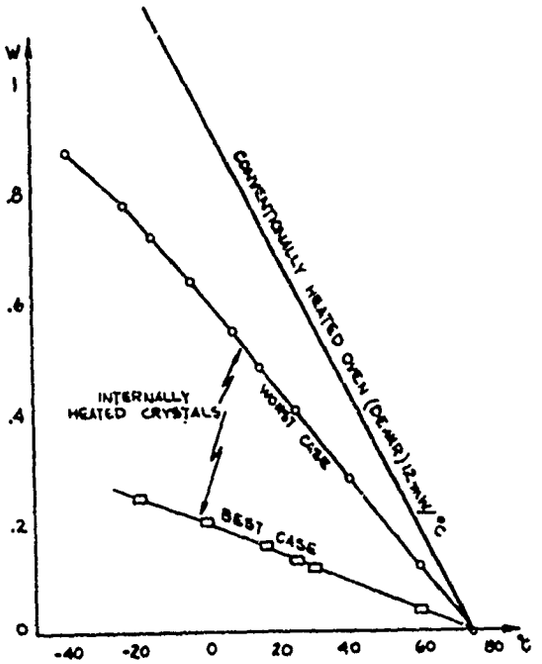


Fig. 70 - Power and transient characteristics of directly heated crystal plates.

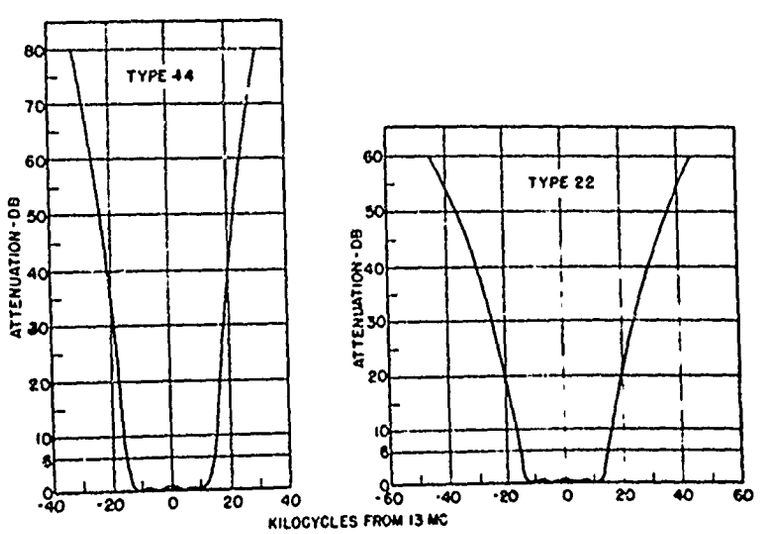


Fig. 71 - 13 MHz crystal filter characteristics.

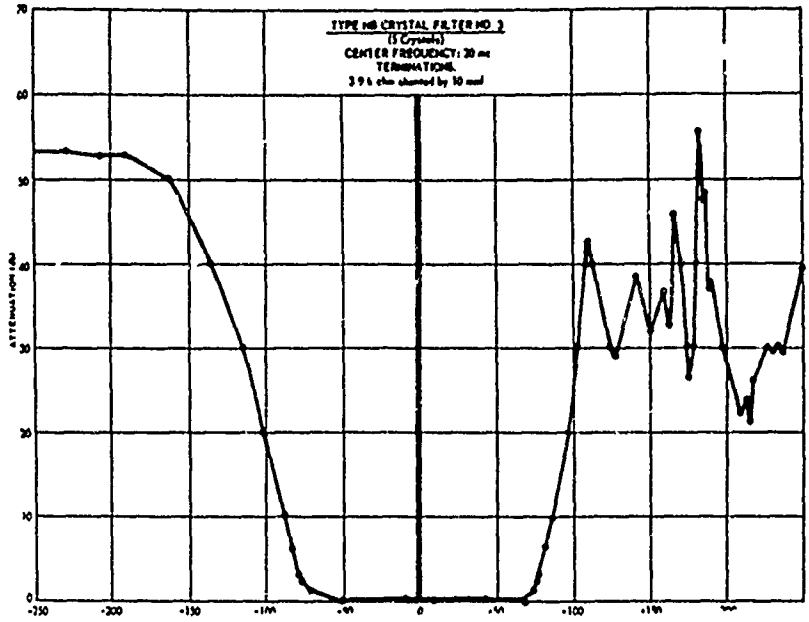


Fig. 72 - Side band crystal filter.

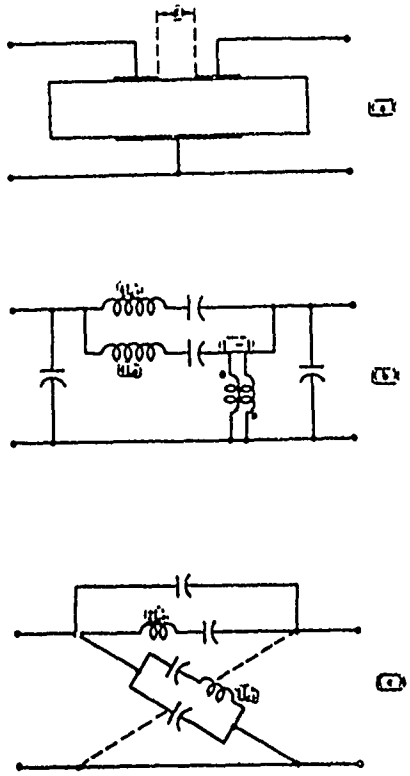


Fig. 73 - Dual resonant crystal and equivalent circuits.

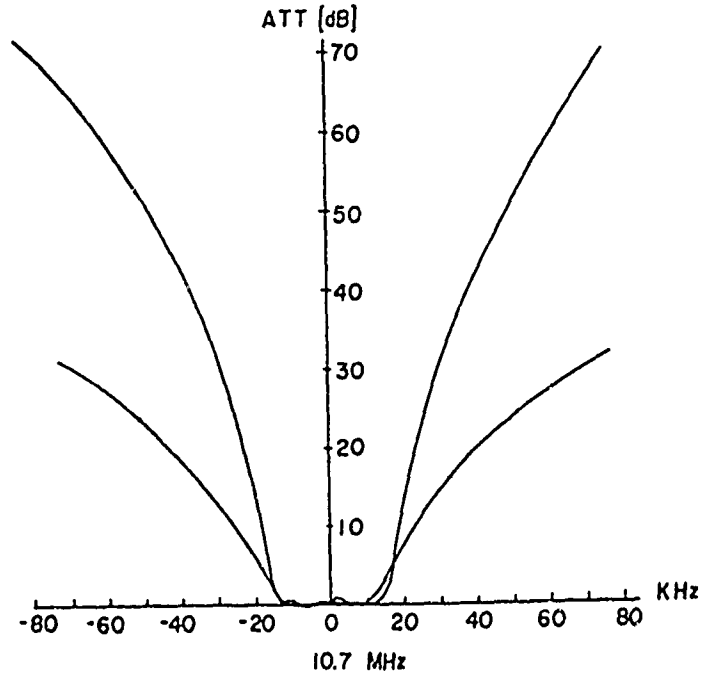


Fig. 74 - Filter characteristics of single and tandem dual filters.

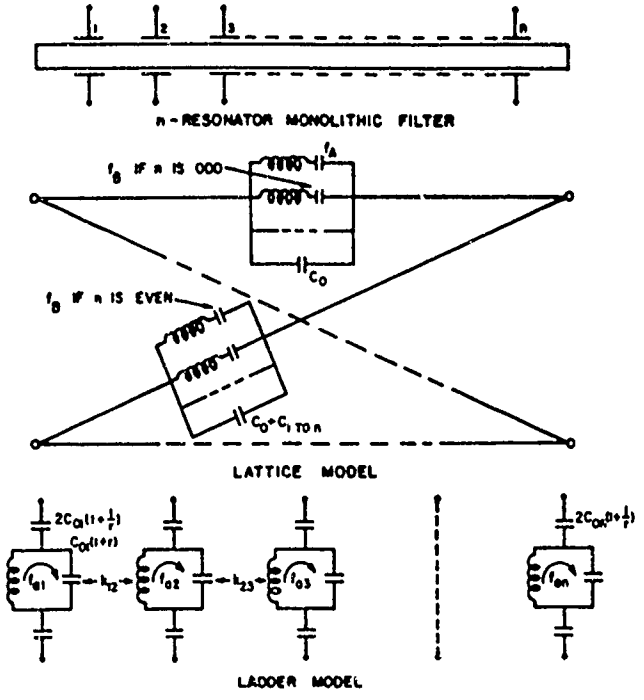


Fig. 75 - Generalized lattice and ladder equivalent of multiresonator crystal filters.

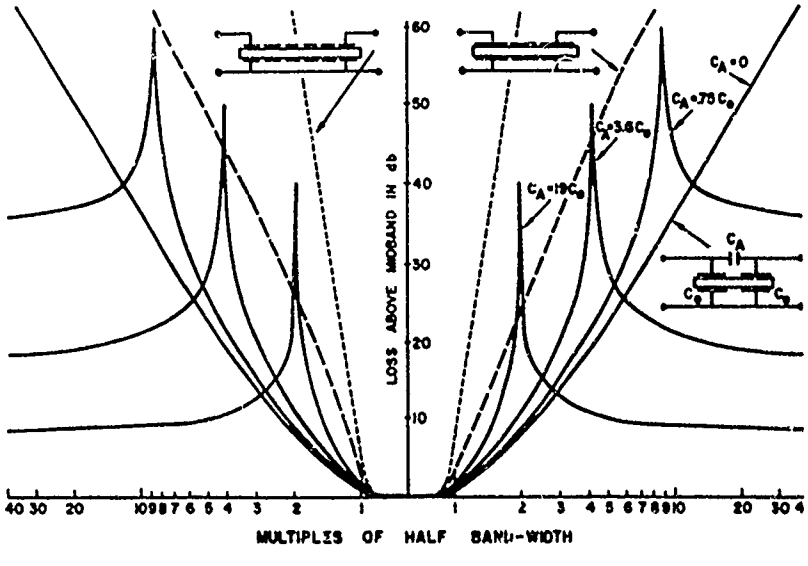
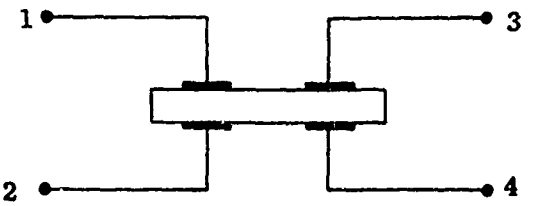


Fig. 76 - Normalized transmission characteristic of monolithic crystal filters.



DUAL RESONATOR CRYSTAL

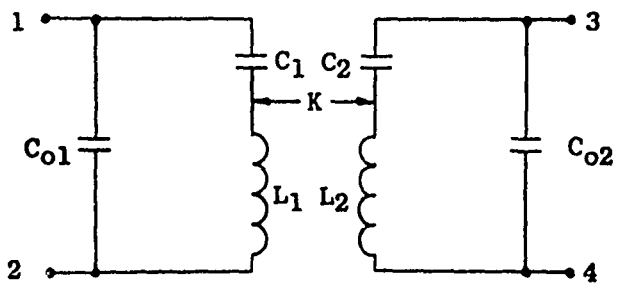


Fig.77- EQUIVALENT CIRCUIT OF DUAL RESONATOR CRYSTAL

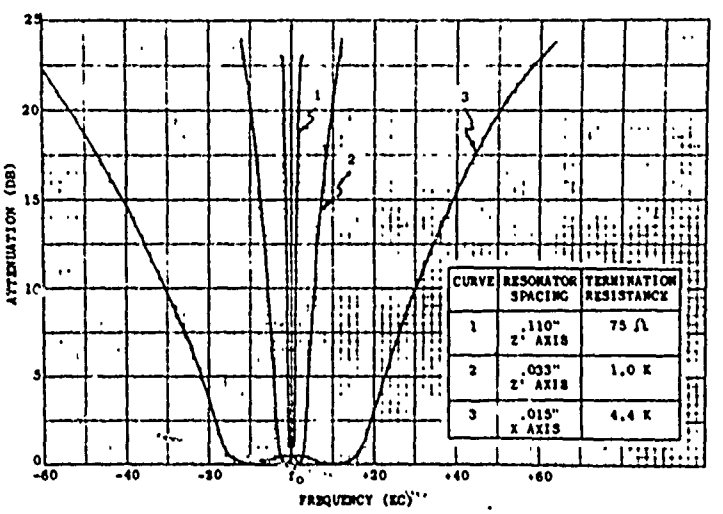


Fig. 78 - Attenuation characteristics of one section electromechanical filters.

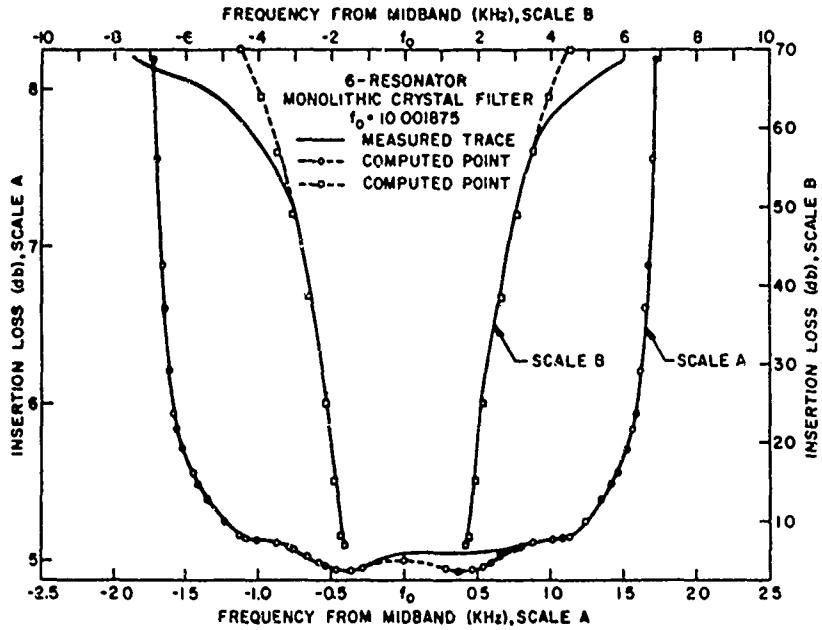
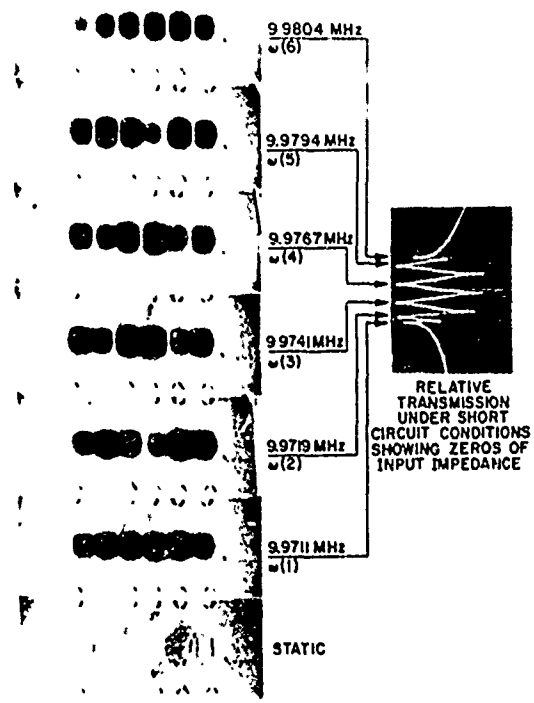


Fig. 79 - 6-resonator monolithic crystal filter.

Fig. 80 - X-ray topographs of critical frequencies of a 6-resonator filter.



MAXIMUM DARKENING IN THE TOPOGRAPHS CORRESPONDS TO THE MAXIMUM AMPLITUDE OF THE ACOUSTICAL DISPLACEMENT THE LIGHT AREAS ARE QUIESCENT OR NODAL REGIONS THE DARK LINES IN THESE LIGHT REGIONS ARE DEFECTS IN THE QUARTZ

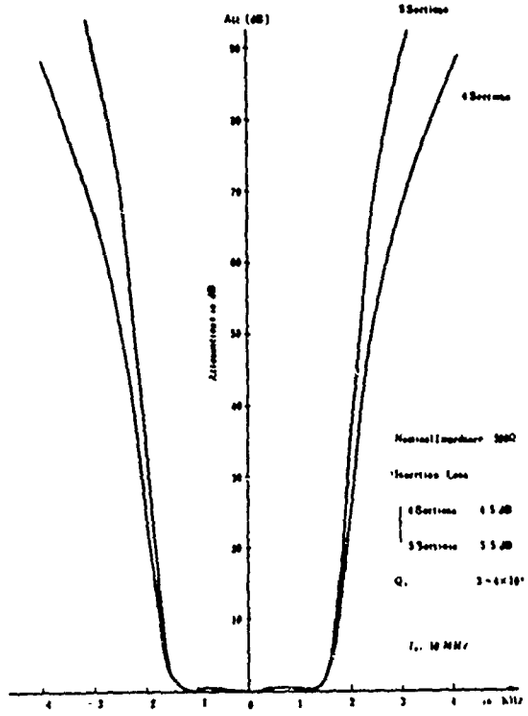


Fig. 81 - SSB narrow band HCM filter.

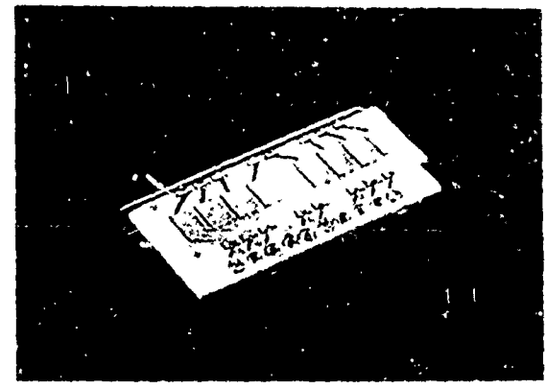


Fig. 83 - 8-section monolithic filter; average and standard deviation of 24 out-of-band characteristics.

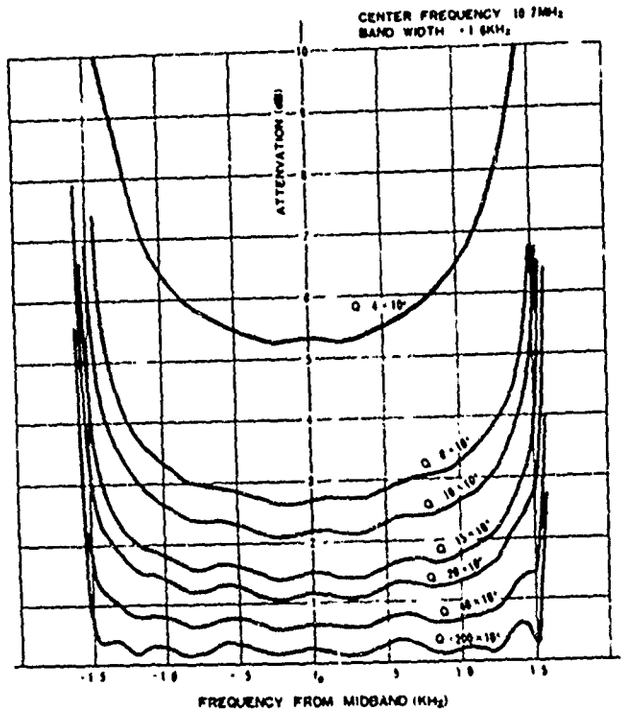
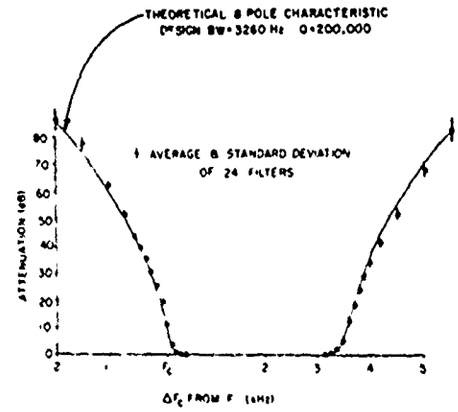


Fig. 82 - Computed inband loss of monolithic filter.



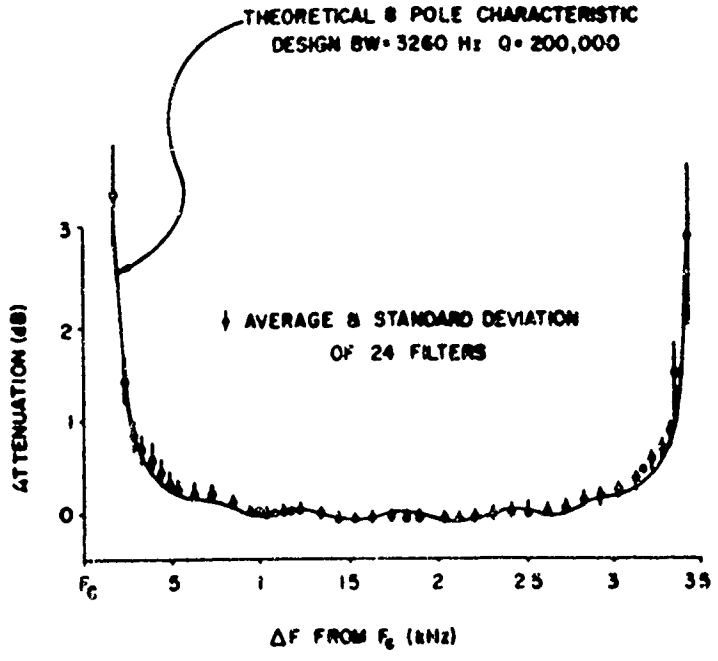


Fig. 84 - Average and standard deviation of 24 MCF in-band characteristics.

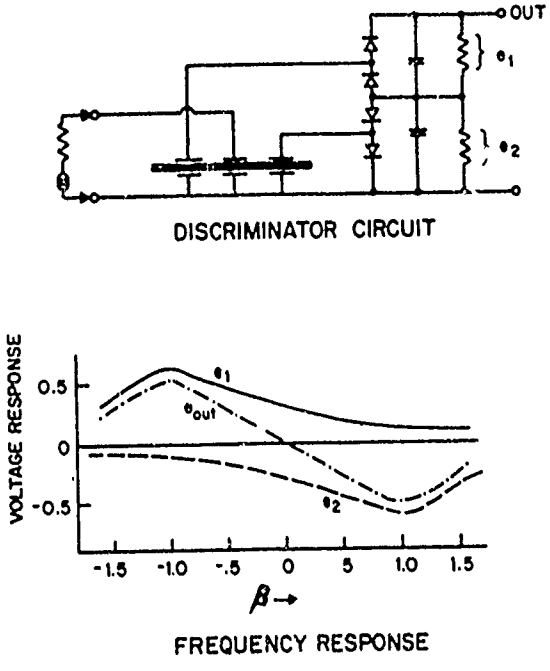
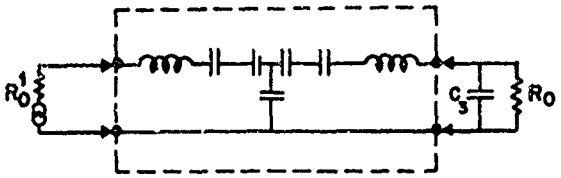


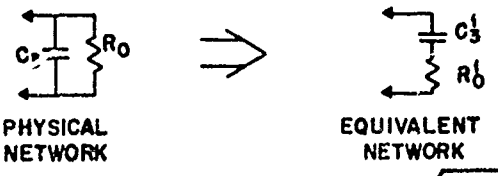
Fig. 85 - Discriminator circuit and response of 3 resonator array.



IMPEDANCE TRANSFORMING NETWORK

FILTER DESIGN IMPEDANCE = R_0'

TRANSFORMATION RATIO = $n = \frac{R_0}{R_0'}$



REQUIRED PHYSICAL CAPACITOR $C_3 = \frac{1}{W_0 R_0}\sqrt{\frac{R_0 - R_0'}{R_0'}}$

EQUIVALENT TUNING CAPACITY $C_3' = \frac{1}{W_0\sqrt{R_0'(R_0 - R_0')}}$

Fig. 86 - Impedance transformation of dual resonators.

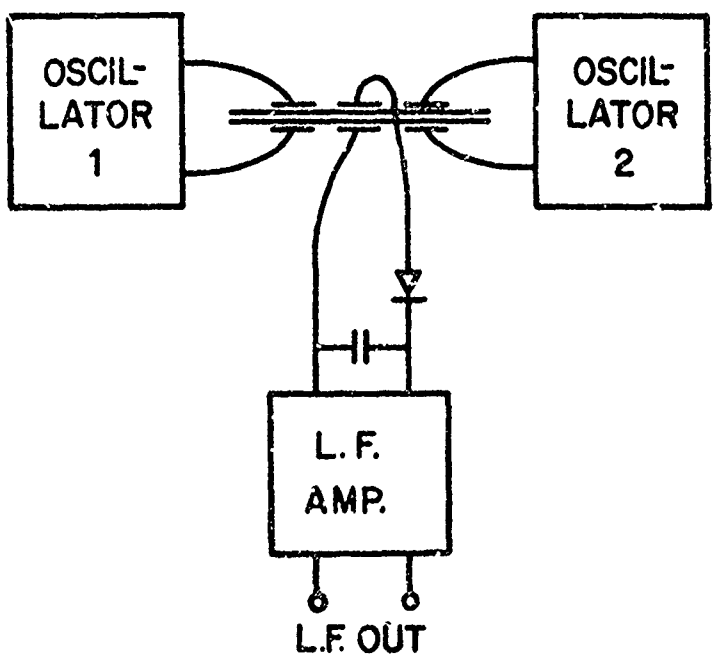


Fig. 87 - Low frequency generator using 3 resonator array.

HISTORY OF ATOMIC AND MOLECULAR CONTROL OF FREQUENCY AND TIME*

Norman F. Ramsey
Harvard University
Cambridge, Massachusetts

Summary

The history of atomic and molecular frequency control will be traced from the earliest molecular beam magnetic resonance experiments in 1938 through the present.

Introduction

In discussing the history of atomic and molecular frequency control two alternative approaches are available. One is to treat all devices in parallel on a year-by-year basis. The other is to discuss each alternative device in succession. It seems clear that the latter approach is the best in the present case so I shall adopt it but I shall make frequent cross references to other devices. In following this procedure, it is clear that the first method discussed should be the molecular and atomic beam magnetic resonance method, not only because historically it was the first important method but also because it stimulated the invention of the other methods and still remains one of the most effective techniques of frequency control. Since the atomic beam devices and the hydrogen masers at present provide the most precise frequency and time standards, the greatest attention will be paid to them, but the history of the other atomic and molecular frequency control devices will be given as well.

Early History of the Molecular Beam Resonance Method

The molecular beam magnetic resonance method arose from a succession of ideas of which the earliest can be traced back to 1927, although it was rather remote from the idea of resonance. In 1927 the physicist, Sir Charles Darwin[1] -- the grandson of the great evolutionist -- discussed theoretically the non-adiabatic transitions which make it possible for an atom's angular momentum components along the direction of a magnetic field to be integral multiples of $\hbar$ both before and after the direction of the field is changed an arbitrary amount. Inspired by Darwin's theoretical discussion, Phipps and Stern[2] in 1931 performed the first experiments on paramagnetic atoms passing through weak magnetic fields whose directions varied rapidly in space. Guttinger[3] and Majorana[4] developed further the theory of such experiments. Frisch and Segrè[5] continued atomic beam experiments with adiabatic and non-adiabatic transitions of paramagnetic atoms and found, in agreement with Guttinger's and Majorana's theories, that transitions took place when the rate of change of the direction of the field was large or comparable to the Larmor frequency,

$$\omega_o = \gamma_I H_o , \qquad (1)$$

which is the classical frequency of precession of a classical magnetized top with the same ratio γ_I of magnetic moment to angular momentum. Transitions did not take place when the rate of change of the direction of H was small compared to the Larmor frequency. However, some of the results of Frisch and Segrè were not consistent with theoretical expectations. Rabi[6] pointed out that this discrepancy arose from the effects of the nuclear magnetic moments since some of the transitions were performed in such weak fields that strong or intermediate coupling between the nuclei and the electrons prevailed. The transitions in such circumstances were quite different from those for which the effects of the nuclear spins could be neglected. Rabi showed that the results of Frisch and Segrè were consistent with expectations if the effects of the nuclei were included. Rabi also pointed out that such non-adiabatic transitions could be used to identify the states and hence to determine the signs of the nuclear magnetic moments. Motz and Rose,[7] Rabi[8] and Schwinger[9] in 1937 calculated the transition probability for molecules which passed through a region in which the direction of the field varied rapidly.

In all of the above experiments, however, the direction of the field varied in space and the only time variation arose as the atoms in the atomic beam passed through the region. Since the atoms possessed a Maxwellian velocity distribution, the atomic velocities varied and the apparent frequencies of the changing field were different for different velocities. Furthermore, the change in field direction ordinarily went through only a portion of a full cycle. For both of these reasons no sharp resonance effects could be expected. No suggestion was made to use an oscillatory magnetic field, i.e., a field that varied in time instead of space so the apparent frequency would be the same to all the atoms, independent of their velocities. It is surprising that this possibility was not immediately recognized after Rabi's brilliant theoretical paper[8] in 1937. To simplify the theoretical analysis Rabi assumed in 1937 that the field was actually oscillatory in time. As a consequence the results are all applicable to the resonance case with oscillatory magnetic fields even though the possibility of actually using fields oscillatory in time was not then recognized. Consequently this paper without alteration still provides the fundamental theory for molecular beam magnetic resonance experiments with oscillatory fields, even though the oscillatory field method was invented by Rabi[10,11] only a year or so after the fundamental theoretical paper was written. Additional amusing characteristics of this brilliant theoretical paper by Rabi[8] are the absence of a verb in the opening sentence and the repeated appearance of $(\nu^2 + \nu^2)$ in the equations when the intended expression was $(\nu^2 + r^2)$.

Gorter[12] in 1936 had suggested that nuclear transitions in solids be induced by an oscillatory field from a radiofrequency oscillator. He proposed to detect the transitions by the absorption of the radiofrequency radiation and by the rise in temperature of solids subject to such oscillatory fields. Although Purcell, Torrey

and Pound[13] and Bloch, Hansen and Packard[14] in 1946 successfully detected the absorption of such transitions by the reaction of the radiation on the radiofrequency circuits, Gorter's experiments[12] were unsuccessful in 1936.

Following a visit by Gorter to Columbia University in September 1937 in which he described his unsuccessful experiments, Rabi[10,11] proposed the use of an oscillator driven magnetic field as the transition inducing field in a molecular beam resonance experiment. Two successful molecular beam devices using this method were soon constructed by Rabi,[10,11] Zacharias,[10,11] Kusch,[10] Kellogg[11] and Ramsey.[11] A schematic view of the first[10] of these is shown in Fig. 1. In these experiments the atoms and molecules were deflected by a first inhomogeneous magnetic field and refocussed by a second one. If a resonance transition were induced in the region between the two inhomogeneous fields, its occurrence could easily be recognized by the reduction of intensity associated with the accompanying failure of refocussing. For transitions induced by the radiofrequency field, the apparent frequency was almost the same for all molecules independent of molecular velocity. As a result, sharp resonances were obtained whenever Eq. (1) was satisfied.

Rabi, Kellog, Zacharias and Ramsey[11] soon extended the method to the molecule H_2 for which the resonance frequencies depended not only on Eq. (1) but also on internal interactions within the molecule. The transitions in this case occurred whenever the oscillatory field was at a Bohr frequency for an allowed transition

$$h\nu = E_1 - E_2 \qquad (2)$$

For the first time they began speaking of their results as "radiofrequency spectroscopy."

Molecular Beam Magnetic Resonance Experiments

By 1939, the new molecular beam magnetic resonance method had demonstrated its usefulness sufficiently well that it appeared to Rabi, Kellogg, Ramsey, and Zacharias to be of possible value for the definition of standard magnetic fields and for use as a time and frequency standard. In 1939 they discussed these possibilities with a scientist at the Bureau of Standards -- whose name is fortunately no longer remembered -- and found little interest in the use of subtle molecular beam techniques for such practical purposes as standards of magnetic field, time or frequency.

In most respects the molecular beam technique in 1939 was more suitable as a standard of magnetic field than of frequency or time since the observed resonances at that time were largely dependent on the externally applied magnetic field. From the point of view of frequency control, it was consequently a great step forward when in 1940 Kusch, Millman and Rabi[15,16] first extended the method to paramagnetic atoms and in particular to $\Delta F = \pm 1$ transitions of atoms where the relative orientation of the nuclear and electronic magnetic moments were changed, in which case the resonance frequencies were determined dominantly by fixed internal properties of the atom rather than by interactions with an externally applied magnetic field. The first resonance measurements of the Cs hyperfine separation, which has been so extensively used in frequency control, were reported[16] in 1940.

In 1941 the research with the atomic beam magnetic resonance method was mostly interrupted by World War II and did not resume before 1946. In 1940 Kusch and Taub[17] pointed out the possibility of observing the hyperfine resonances at magnetic fields such that the resonance frequency was an extremum, in which case the frequency to first order was independent of the strength of the magnetic field.

In 1949 Ramsey[18,19] invented the separated oscillatory field method for molecular beam resonance for which the oscillatory field, instead of being distributed uniformally throughout the transition region, was concentrated in two coherently driven oscillatory fields in short regions at the beginning and end of the transition region. The theoretical shape of a resonance curve with this apparatus is shown in Fig. 2. Ramsey pointed out that this method has the following advantages: (a) The resonances are 40% narrower than even the most favorable Rabi resonances with the same length of apparatus; (b) The resonances are not broadened by field inhomogeneities; (c) The length of the transition region can be much longer than the wave length of the radiation provided the two oscillatory regions are short whereas there are difficulties with the Rabi method due to phase shifts when the length of the oscillatory region is comparable to the wave length; (d) The first order Doppler shift can mostly be eliminated when sufficiently short oscillatory field regions are used; (e) The sensitivity of resonance measurements can be increased by the deliberate use of appropriate relative phase shifts between the two oscillatory fields. All of these characteristics are of great value for the use of atomic beam resonance devices equipment as precision frequency and time standards. A photograph of an early molecular beam apparatus[20] using this method is shown in Fig. 3.

Atomic Beam Frequency Standards

With the above developments, it was apparent to most molecular beamists by 1949 that atomic beam methods could be highly effective for precision frequency control. However, this was less clear to others who then believed that the advances in crystal frequency control techniques had been so great that atomic devices could not be enough better to justify the extra cost and effort. However, in 1952 Sherwood, Lyons, McCracken and Kusch[21,22] reported briefly on atomic beam resonance research supported by the National Bureau of Standards and directed primarily toward the development of an atomic beam clock. A schematic diagram of a proposed atomic beam clock at that time is given in Fig. 4. The financial support for such work soon dwindled due to advances in microwave spectroscopy and the view then held by those supporting the work that a molecular clock based on the microwave absorption by ammonia at its inversion frequency would be simpler and more promising.

A few years later Zacharias[23,24] stimulated renewed interest in an atomic beam cesium clock. His initial concern was for an entirely new type of cesium beam in which ultra high precision would be obtained by the use of extremely slow molecules moving upwards in a vertical apparatus at such low velocities that they would fall back down by the action of gravity. Although this fountain experiment eventually failed due to the unexpectedly great deficiency of the required ultra slow molecules emerging from the source, it stimulated Zacharias to develop and to urge others to develop well engineered atomic beam frequency standards using normal atomic velocities. The fountain experiment of Zacharias illustrates the value to science of even some unsuccessful experiments since the existence of this unsuccessful effort directly and indirectly stimulated three quite different developments: (a) The use of conventional but well engineered atomic beams for frequency control; (b) The development by Ramsey and others[12] of the stored atom technique which eventually lead to the hydrogen maser and (c) High preci-

sion resonance experiments with ultra slow neutrons.[43] The first report on an atomic beam frequency standard at a Frequency Control Symposium was that of Zacharias at the Ninth Symposium in 1955 (the first report of any kind on atomic or molecular research at a Frequency Symposium was a paper at the Fourth Symposium in 1950 by R. H. Dicke[26] on the reduction of Doppler widths in microwave absorption). Zacharias claimed a short time stability of 1 part in 10^9 for his atomic cesium frequency standard.

In 1955, Essen and Parry[25] of the British National Physical Laboratory successfully operated the first practical laboratory atomic cesium beam apparatus that was extensively used as an actual frequency standard. Their construction and effective use of this device provided a major impetus to the subsequent development of atomic beam cesium frequency standards.

In 1956 the first commercial model of an atomic beam frequency standard appeared on the market. This was the National Company's "Atomichron" developed[27] by J. H. Holloway and A. Orenberg in collaboration with J. R. Zacharias and later joined by R. McCoubrey and R. Daley. This device used Ramsey's separated oscillatory field method for increased precision, a special design of cesium oven which could be operated several years without exhaustion, titanium pumping to permit permanent sealing off of the evacuated beam tube, and many other features necessary for an effective commercial device. A photograph of the first commercial Atomichron is shown in Fig. 5. The development of the Atomichron was largely supported financially by the U. S. Signal Corps at Ft. Monmouth though some support came from the Air Force. A purchase order by the Signal Corps for a relatively large number of Atomichrons made possible the development of mass production techniques and improved engineering to permit sufficient reliability and reductions in price to assure commercial success.

The early atomic beam frequency standards were subject to various frequency shifts dependent on the amplitude of the radiofrequency power used and on other variables. To account for these results Ramsey analyzed, with the aid of a computer analysis supported by the National Company, the various possible distortions that would occur in an atomic beam resonance.[28] The elimination of radiofrequency phase shifts and other sources of distortion made possible the marked increases in accuracy that have been obtained with the atomic beam frequency standards.

From 1956 on the atomic beam frequency standards developed rapidly. Mockler, Beeler and Barnes[26,27] developed an atomic cesium frequency standard at the National Bureau of Standards in Boulder. Other commercial organizations such as TRG, Bomac, Varian, and Hewlett-Packard became involved. Many laboratories outside of both England and the United States either constructed or purchased atomic beam frequency standards including those in Canada, France, and Germany and the laboratories of Kartaschoff[26] and Bonanomi[26] in Switzerland. Other materials than cesium were tested for molecular beam clocks; although thallium and various polyatomic molecules have some significant advantages they have not yet significantly displaced cesium which is particularly convenient. Reder, Winkler and others[26] at Ft. Monmouth and Markowitz at the Naval Observatory sponsored various world-wide studies of the comparison of atomic clock frequencies and the synchronization of clocks. Extensive studies were made of other atoms such as thallium for use in the atomic beam tubes and various molecular resonances were studied for possible use in a molecular beam electric resonance apparatus for frequency control purposes. However, atomic cesium remains the most widely used substance in molecular or atomic beam frequency control devices.

Particularly effective atomic beam cesium clocks were developed and sold by Hewlett-Packard which also developed a "flying clock" particularly suitable for the intercomparison of atomic clocks in different laboratories. A modern beam tube for an atomic cesium frequency standard is shown photographically in Fig. 6.

In 1967, the Thirteenth General Conference of Weights and Measures resolved that the unit of time in the International System of Units should be the second defined as follows: "The second is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium atom 133."

Microwave Absorption Spectroscopy

Microwave absorption spectroscopy had an early start in the experiments of Cleeton and Williams[29,30] in 1934. They observed the absorption of microwave radiation at the NH_3 inversion frequency. However, research on microwave absorption was inhibited at that time by the lack of suitable microwave oscillators and circuits so there was no further development of microwave absorption spectroscopy until after the development of microwave oscillators and wave guides for radar components in World War II. Immediately following World War II there was a great burst of activity in microwave absorption spectroscopy. Although there were no publications on experimental microwave spectroscopy in 1945, in the single year of 1946 there were a number of important publications from many different laboratories including reports by the following authors[31]: Bleaney, Penrose, Beringer, Townes, Dicke, Strandberg, Dailey, Kyhl, Van Vleck, Wilson, Dakin, Good, Coles, Hershberger, Lamont, Watson, Roberts, Beers, Hill, Merritt and Walter, and in 1947 there were over 60 published papers on this subject including a number of publications by Gordy and Jen as well as those with reports the previous year and by others. A typical microwave absorption experiment at this time is shown schematically in Fig. 7.

Microwave absorption techniques were quickly recognized to be of potential value for frequency standards. In 1948 a group of workers[22] at the National Bureau of Standards built an ammonia clock which was completed in 1949 and is shown in Fig. 8 and which eventually achieved an accuracy of a part in 10^8. J. Rossel[22] in Switzerland and K. Shimoda in Japan devised an improved ammonia absorption clock good to a few parts in 10^9.

The first report at a Frequency Control Symposium pertaining to atomic and molecular frequency standards was that of Dicke at the Fifth Symposium in 1951 when he delivered a report entitled, "Reduction of the Doppler Contribution to the Width of Microwave Absorption Lines." In the Seventh, Eighth and Ninth Symposia he and Carver, Arditi, and others described the continuation of this work at both Princeton and R.C.A. with the financial support of the Signal Corps. The microwave absorption studies soon merged with the optical pumping techniques described in the next section, since the intensities of the resonances were greatly enhanced by the use of optical pumping.

Optical Pumping

The starting point of all research optical pumping was a paper by Bitter[32] in 1949 which showed the possibility of studying nuclear properties in optically excited states. Kastler[33,34] showed the following year that this technique could be effectively combined with the double resonance method he and Brossel[33] had developed. Both optical pumping and optical detection techniques serve the purpose of increasing the signal to noise ratio of the resonator

output signal: the optical pumping greatly enhances the population of certain states so the signal is not weakened by stimulated emission nearly cancelling absorption, and the optical detection increases the signal to noise ratio because of the lower noise level of optical detectors over microwave detectors.

The combination of optical pumping techniques with the buffer gas method for reducing Doppler shift developed by Dicke[27] and others[27] provided gas cells of real value as frequency control devices. Although many different atoms have been used in such gas cells, Rb^{87} soon became the favorite in most such devices. Extensive work in optically pumped gas cells for frequency control has been done at Princeton, R.C.A., I.T.T., Space Technology Laboratory, the National Bureau of Standards, Clauser Technology Corporation, Varian Associates, and many other commercial, university and government organizations in the U. S. and abroad. Fig. 9 shows a typical optically pumped rubidium frequency standard.

The optically pumped gas cells have the advantages of simplicity, relatively low cost, large signal to noise ratio, and good spectral purity. Unfortunately the relatively large shift in frequency due to numerous buffer gas collisions is dependent on purity, pressure and temperature. As a result, the stability of rubidium gas cells over a period of several months is ordinarily no better than a few parts in 10^{10}. These pressure shifts prevent the optically pumped gas cells from being primary frequency standards but the gas cells are used as frequency control devices when too much accuracy is not required. Research is currently in progress in a number of laboratories to improve the stability of optically pumped gas cells; Bouchiat,[35] Brossel[35] and others, for example, have eliminated the buffer gases and as in the hydrogen maser have used many fewer collisions with suitably coated walls to retain the atoms and reduce the effect of second order Doppler shift.

Molecular Masers

In 1951 Pound, Purcell and Ramsey[36] studied nuclear spin systems with inverted populations and noted that such systems in principle were intrinsic amplifiers rather than absorbers. The first suggestions actually to use systems with inverted populations as practical amplifiers and oscillators were made at closely the same time in 1953-1955 and independently by Townes,[37] by Weber[38] and by Basov and Prokhorov.[39] The first such amplifier was successfully constructed in 1955 by Gordon, Zeiger, and Townes[1] and called a MASER (Microwave Amplifier by Stimulated Emission of Radiation). The device used inhomogeneous electric fields to focus the higher energy molecular inversion states of ammonia molecules in a molecular beam. These molecules then emitted coherent stimulated radiation in passing through a cavity tuned to the 24 GHz ammonia inversion transition. A schematic diagram of the first ammonia maser is shown in Fig. 10. A report by Gordon on the new ammonia maser was a major attraction at the special meeting on atomic and molecular resonances sponsored by the Signal Corp. Engineering Laboratory in 1956. In that year Bloembergen[40] proposed the three level solid state maser and in 1958 Townes and Schawlow[41] pointed out the possibility of masers at the infrared and optical frequencies.

Since the announcement of the first successful ammonia maser in 1955 there has been tremendous research and development activity by scientists and engineers in many countries. Masers at infrared or optical frequencies (often called lasers) have great potential for frequency control even though they have not been used for this purpose as yet. Further discussion of lasers will be deferred to a later section of this report. Molecular maser developments for the purposes of frequency control soon became intense and went in many directions including the search for more suitable molecules than ammonia, the development of two cavity masers (analogous to the separated oscillatory field method[18] for molecular beams), use of ammonia of different isotopic composition, etc. However, after a few years of intense molecular maser activity, the interest in such masers for frequency control waned since the molecular masers on the one hand lacked the simplicity and low cost of optically pumped rubidium gas cells, and on the other hand lacked the high precision of either atomic cesium beams or atomic hydrogen masers.

Atomic Masers

In 1957 Ramsey[26] proposed to increase the accuracy of the atomic beam magnetic resonance method by retaining the atoms for a much longer time between the two separated oscillatory fields, thereby obtaining much narrower resonances. His first thought was to confine the atoms with inhomogeneous magnetic fields in a large ring such as the circular tunnel of the 6 GeV Cambridge Electron Accelerator. However, it soon became apparent that the inhomogeneous confining magnetic fields which acted on the atoms for long periods of time, would hopelessly broaden the resonances. In fact, it became clear that the frequencies would be much less perturbed by a confinement force that was present for only a short fraction of the time even though the force might be stronger when it was applied. The obvious limit of such a device was confinement of atoms in a box with suitably coated walls. Although many wall bounces would be required to achieve marked narrowing of the resonance by long storage time, the first experiments involved only a few wall collisions, since most scientists at that time believed that even atoms in an S state would undergo hyperfine transitions at even a single wall collision. The first experiments of Kleppner, Ramsey and Fjelstadt[42] involved only a few wall collisions and the experiment was appropriately called a "broken atomic beam resonance experiment." Cesium atoms and a teflon coated wall were used in these first experiments.

Goldenberg, Kleppner and Ramsey[44] then made an atomic beam resonance apparatus which stored atoms of cesium for a longer time and investigated alternative wall coating materials. They found that when the storage bulb was coated with a paraffin-like substance called Paraflint[44] resonances could be observed after as many as 200 wall collisions. It was recognized that atomic hydrogen would probably be a more suitable atom than cesium because of the low electric polarizability and the low mass of hydrogen, but cesium could be much more efficiently detected than hydrogen.

Kleppner and Ramsey[43,44,45] therefore proposed detection of the emitted radiation rather than of the atom. In particular, they noted that atoms of hydrogen in the higher energy hyperfine state could be focussed into a suitably coated storage bulb by a six pole magnet while atoms in the lower state would be defocussed. They showed that if such a storage bulb were surrounded by a microwave cavity tuned to the 1420 MHz hyperfine transition frequency, then maser oscillation should occur. In 1960, Goldenberg, Kleppner, and Ramsey[46] constructed and operated the first atomic hydrogen maser. A photograph of this apparatus is shown in Fig. 11. Although the total microwave power was small -- approximately 10^{-12} watts -- the stability was so high that the output was concentrated into an extremely narrow band with a consequently favorable signal to noise ratio.

Although the first hydrogen masers used wall coatings of Paraflint[44] (a variety of paraffin) or of Dri-Film[44] (dimethyldichlorosilane), it was soon found that with atomic hydrogen, in contrast to cesium, teflon coated walls gave longer storage times and smaller frequency shifts from wall collisions.[46] Bender[47] soon pointed out that spin exchange collisions of hydrogen atoms could not be neglected and might produce a significant frequency shift but Crampton[48] noted that the normal tuning technique would cancel out such an effect.

A commercial hydrogen maser was developed by Vessot, Peters, Vanier, McCoubrey, Levine, and Cutler.[49] The work was started at Bomac and successively transferred to Varian Associates and Hewlett-Packard. It is currently being carried on by Vessot and his associates at the Smithsonian Astrophysical Observatory and at the Jet Propulsion Laboratory. The H-10 maser developed by Vessot and his associates is shown in Fig. 12 and 13. The masers are being built chiefly for long base line interferometry in radio astronomy, which benefits greatly from the high stability of the hydrogen maser. Research and development on hydrogen masers is carried out in the laboratories of Kartaschoff[26] in Switzerland, Audoin[58] and Grivet[26] in France, and in a number of other countries.[59]

The chief disadvantage of the hydrogen maser for time and frequency control has been the existence of a small frequency shift due to collisions of the atoms with the teflon coated walls of the storage bulb. With a 16 cm. diameter bulb this wall shift is about two parts in 10^{11} and can be measured by using bulbs of two different diameters. However, until recently the measurements of the wall shifts have been limited to accuracies of a few percent by variations in different wall coatings. However, Uzgiris and Ramsey[50] at Harvard have reduced the wall shift by a factor of ten by the use of an atom storage vessel of ten times larger diameter (1.5m). In the same laboratory, Brenner[51] and Debely[52] have recently developed a technique to measure the wall shift in a single storage bulb by changing its volume by deforming its shape. Since a single bulb is used in this method, it is free from the uncertainties in the non-reproducibility of the wall coatings of different bulbs. Although this method has so far been usable on only hydrogen masers with normal sized storage bulbs, it should also be applicable to the large storage bulbs. Zitzewitz[53] has also shown that at a temperature of about 80°C the wall shift passes through zero; it is thus possible to operate the hydrogen maser at a temperature such that the wall shift vanishes and to select this temperature by the deformable bulb technique. With these new methods, it is hoped that an absolute accuracy considerably better than 1 part in 10^{13} can be attained.

Although the hydrogen maser is the most stable atomic maser for long periods of time, Novick,[26] Vanier[26] and others have developed a high power optically-pumped atomic Rb^{85} maser whose relatively high output power is useful for short term stability.

Optical Masers

Townes and Schawlow[41] pointed out that masers could be produced at infrared and optical frequencies. The first optical maser or laser was successfully made from ruby by Maiman.[54] Subsequently there was a great burst of activity in this field and lasers were made of a wide variety of materials and at high pulsed power. From the point of view of frequency control, the laser using a helium-neon gas mixture developed by Javan[55] and his associates is of particular interest because of its potentially greater stability. As absolute time standards, most lasers suffer from the most sensitive determination of the output frequency being the distance between two mirrors while in the the atomic hydrogen maser, the primary determination of frequency is the hydrogen atom itself with only a relatively small amount of pulling from mistuning of the microwave cavity. However, various methods for determining the laser frequency more by the atomic properties have been developed and further improvements are being attempted. Likewise Javan[56] and others are making vigorous efforts by frequency multiplication to bridge the gap between microwave radiation and the optical laser frequency. When this is done the lasers may provide a valuable alternate in frequency control. Since this has not been accomplished so far, lasers have not yet been used for frequency control and they are consequently discussed only briefly in this review of frequency control devices.

Trapped Ions

Dehmelt[56] in 1959 first used electromagnetic ion traps in radiofrequency resonance studies. Since the ions can be retained in the apparatus for very long times, the resonances could be very narrow by the uncertainty principle. However, so far the resonances have been significantly broadened by the second order Doppler shift arising from the ion velocities in the ion trap. Efforts have been made to diminish the broadening by Dehmelt, Major, Fortson and Schuessler[57] but such trapped ion devices have not yet provided as stable frequencies as the best alternative frequency standards.

Future Prospects

Although atomic and molecular frequency control has been a reality for a number of years, new developments are still occurring at a relatively rapid rate. As a result it is impossible to forecast reliably the future developments that will lead to the most major subsequent advances. For highest stability and reproducibility the most promising prospects now appear to be (a) further improvements in the atomic and molecular beam methods, (b) hydrogen maser improvements by combinations of the new deformable bulb technique with either the large box maser or operation at a temperature where the wall shift vanishes, (c) improved stored ion devices and (d) use of lasers for frequency control. However, there also may be new ideas and developments that drastically improve one of the existing techniques or lead to totally new methods of atomic or molecular frequency control.

* This paper provided the basis for an invited talk at the 25th Anniversary Frequency Control Symposium on April 26, 1971. Since this is the first effort ever made to present a coherent historical account of this subject, omission and errors are inevitable. The author will welcome any letters either correcting or adding to this record.

1. C. Darwin, Proc. Roy. Soc. 117, 258 (1927).
2. T. E. Phipps and O. Stern, Zeits. of Physik 73, 185 (1931).
3. P. Guttenger, Zeits. J. Physik 73, 169 (1931).
4. E. Majorana, Nuovo Cimento 9, 43 (1932).
5. R. O. Frisch and E. Segrè, Zeits. J. Physik 80, 610 (1933).
6. I. I. Rabi, Phys. Rev. 49, 324 (1936).
7. L. Motz and M. Rose, Phys. Rev. 50, 348 (1936).
8. I. I. Rabi, Phys. Rev. 51, 652 (1937).
9. J. Schwinger, Phys. Rev. 51, 645 (1937).
10. I. I. Rabi, J. R. Zacharias, S. Millman and P. Kusch, Phys. Rev. 53, 318 (1938) and 55, 526 (1939).
11. J. M. B. Kellogg, I. I. Rabi, N. F. Ramsey, and J. R. Zacharias, Phys. Rev. 55, 729 (1939); 56, 728 (1939); and 57, 677 (1940).
12. C. J. Gorter, Physica 3, 503 and 995 (1936).
13. E. M. Purcell, H. C. Torrey and R. V. Pound, Phys. Rev. 69, 37 (1946) and 73, 679 (1948).

14. F. Bloch, W. Hansen and M. E. Packard, Phys. Rev. 69, 127 (1946) and 70, 474 (1946).
15. P. Kusch, S. Millman and I. I. Rabi, Phys. Rev. 57, 765 (1940).
16. S. Millman and P. Kusch, Phys. Rev. 57, 438 (1940).
17. P. Kusch and H. Taub, Phys. Rev. 75, 1477 (1949).
18. N. F. Ramsey, Phys. Rev. 76, 996 (1949) and 78, 695 (1950) and Molecular Beams, Oxford University Press (1956 and 1969).
19. N. F. Ramsey and H. B. Silsbee, Phys. Rev. 84, 506 (1951).
20. H. G. Kolsky, T. E. Phipps, N. F. Ramsey and H. B. Silsbee, Phys. Rev. 80, 483 (1950).
21. J. E. Sherwood, H. Lyons, R. H. McCracken, and P. Kusch, Bull. Am. Phys. Soc. 27, (1), 43 (1952).
22. H. Lyons, Annals of New York Academy of Sciences 55, 831 (1952) and Scientific American 196, No. 2, 71 (Feb. 1957).
23. J. R. Zacharias, private communication, and Phys. Rev. 94, 751 (1954). R. Weiss and R. Vessot were associated with Zacharias in the experimental work on the "fountain" experiment.
24. J. R. Zacharias, J. G. Yates, and R. D. Haun, Mass. Inst. of Tech. Res. Lab. for Electronics, Quarterly Progress Report, January 1955, page 30, and Proc. Inst. Radio Engineers 43, No. 3, 364 (1955).
25. L. Essen and V. L. Parry, Nature 176, 280 and 284 (1955).
26. Programs and Proceedings of the Frequency Control Symposia from 1946 to 1970, IEEE Trans. Instr. Meas. IM-13(1964), IM-15(1966) and to be published, and IEEE Journ. of Quantum Electronics QE5, 43 (1969).
27. F. H. Reder, Atomic Clocks and Their Applications, USASRDL Technical Report 2230 (AD 265452) (1961).
28. N. F. Ramsey, Phys. Rev. 100, 1191 (1955); 109, 822 (1958); Journ. Phys. et Radium 19, 809 (1958); Recent Research in Molecular Beams, Academic Press, New York (1959).
29. C. E. Cleeton and N. H. Williams, Phys. Rev. 45, 234 (1934).
30. E. V. Condon and H. Odishaw, Handbook of Physics, McGraw Hill Book Company, New York (1967).
31. C. H. Townes and A. L. Schowlow, McGraw Hill Book Company, New York (1955).
32. F. Bitter, Phys. Rev. 76, 833 (1949) and M. H. T. Pryce, Phys. Rev. 77, 136 (1950).
33. J. Borssel and A. Kastler, Compt. rend. 229, 1213 (1949).
34. A. Kastler, J. Phys. Radium 11, 255 (1950) and J. Opt. Soc. America 47, 460 (1957).
35. M. A. Bouchiat and J. Brossel, Phys. Rev. 147, 41 (1966).
36. R. V. Pound, E. M. Purcell and N. F. Ramsey, Phys. Rev. 81, 156, 278 and 279 (1951) and 103, 20 (1956).
37. J. P. Gordon, H. J. Zeiger, and C. H. Townes, Columbia Radiation Lab. Progress Report (Dec. 1951); J. Comms. Eng. Japan 36, 650 (1953); Phys. Rev. 95, 282 (1954); and Phys. Rev. 99, 1264 (1955).
38. J. Weber, Transactions of IRE, PG on Electron Devices 3, 1 (1953).
39. N. G. Basov and A. M. Prokhorov, Zh. Eksperim i Teor. Fiz 27, 431 (1954) and 28, 249 (1955) or JETP 1, 184 (1955).
40. N. Bloembergen, Phys. Rev. 104, 324 (1956).
41. A. L. Schowlow and C. H. Townes, Phys. Rev. 112, 1940 (1958).
42. D. Kleppner, N. F. Ramsey, and P. Fjelstadt, Phys. Rev. Letters 1, 232 (1958).
43. J. K. Baird, P. D. Miller, W. Dress and N. F. Ramsey, Phys. Rev. 179, 1285 (1969).
44. H. M. Goldenberg, D. Kleppner and N. F. Ramsey, Phys. Rev. 123, 530 (1961).
45. H. M. Goldenberg, D. Kleppner and N. F. Ramsey, Phys. Rev. Letters 8, 361 (1960).
46. D. Kleppner, H. M. Goldenberg and N. F. Ramsey, Phys. Rev. 126, 603 (1962), and H. C. Berg and D. Kleppner, Rev. Sci. Inst. 33, 248 (1962).
47. P. L. Bender, Phys. Rev. 132, 2154 (1963).
48. S. B. Crampton, Phys. Rev. 158, 57 (1967).
49. D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. C. Vessot, H. E. Peters and J. Vanier, Phys. Rev. 138, A972 (1965).
50. E. Uzgiris and N. F. Ramsey, Phys. Rev. A1, 429 (1970).
51. D. Brenner, Journ. Appl. Phys. 41, 2942 (1970).
52. P. E. Debely, Rev. Sci. Inst. 41, 1290 (1970).
53. P. W. Zitzewitz and N. F. Ramsey, Phys. Rev. A3, 51 (1971).
54. T. H. Maiman, Nature 187, 493 (1960).
55. A. Javan, W. Bennett and D. R. Herriott, Phys. Rev. Letters 6, 106 (1961).
56. H. G. Dehmelt, Phys. Rev. 109, 381 (1959); Advances in Atomic and Molecular Physics 3, 53 (1967), and 5, 109 (1969); and Proc. Intl. Symp. of Physics of One-and-Two Electron Atoms (North Holland Publ. Co., Amsterdam).
57. H. G. Dehmelt, F. Major, E. N. Fortson, and H. A. Schuessler, Phys. Rev. Letters 8, 213 (1967); Phys. Rev. 170, 91 (1968); Phys. Rev. 187, 5 (1969).
58. C. Audoin, Revue de Physique Appliquée, 1, 2 (1966) and 2, 309 (1967) and Phys. Letters 28A, 372 (1968). C. Audoin, M. Desaintfuscien, P. Petit, and J. P. Schermann, Nucl. Instr. and Methods, 69, 1 (1969); IEEE Transactions on Instr. and Meas. IM-17, 351 (1968), (this work utilizes a useful double focussing method to eliminate the undesired F = 1 mF = 1 state from the focussed beam); Electronics Letters 5, No. 13 (1969); C. R. Acad. Sc. Paris 264, 698 (1967) and 270, 906 (1970); IEEE Jour. of Quant. Elect. QE-5, 431 (1969). S. Haroche, C. Cohen-Tanoudji, C. Audoin and J. P. Schermann, Phys. Rev. Letters 24, 861 (1970).
59. Laboratories which have engaged in hydrogen maser studies include Harvard University, Bomac Laboratories, Varian Associates, Hewlett-Packard, National Bureau of Standards, Massachusetts Institute of Technology, Naval Research Laboratory, Goddard Space Flight Center, Jet Propulsion Laboratory, U. S. Army Electronics Command, PTB (Braunschweig, Germany), National Research Council (Canada), R. R. L. (Tokyo, Japan), LSRH (Neuchatel, Switzerland), and the Lebedev Institute (Moscow, USSR).

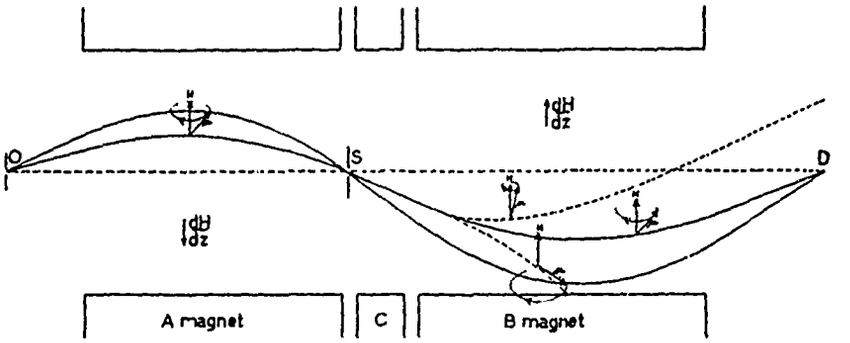


Fig. 1. Schematic diagram[10] showing the principle of the first molecular beam resonance apparatus. The two solid curves indicate two paths of molecules having different orientations which are not changed during passage through the apparatus. The two dashed curves in the region of the B- magnet indicate two paths of molecules whose orientation has been changed in the C- region so the refocussing is lost due to the change in the component of magnetic moment along the direction of the magnetic field.

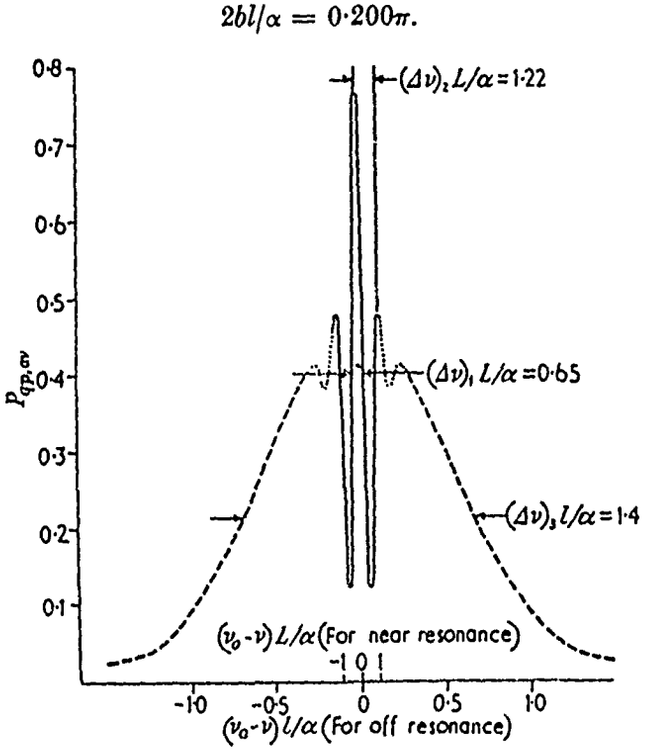
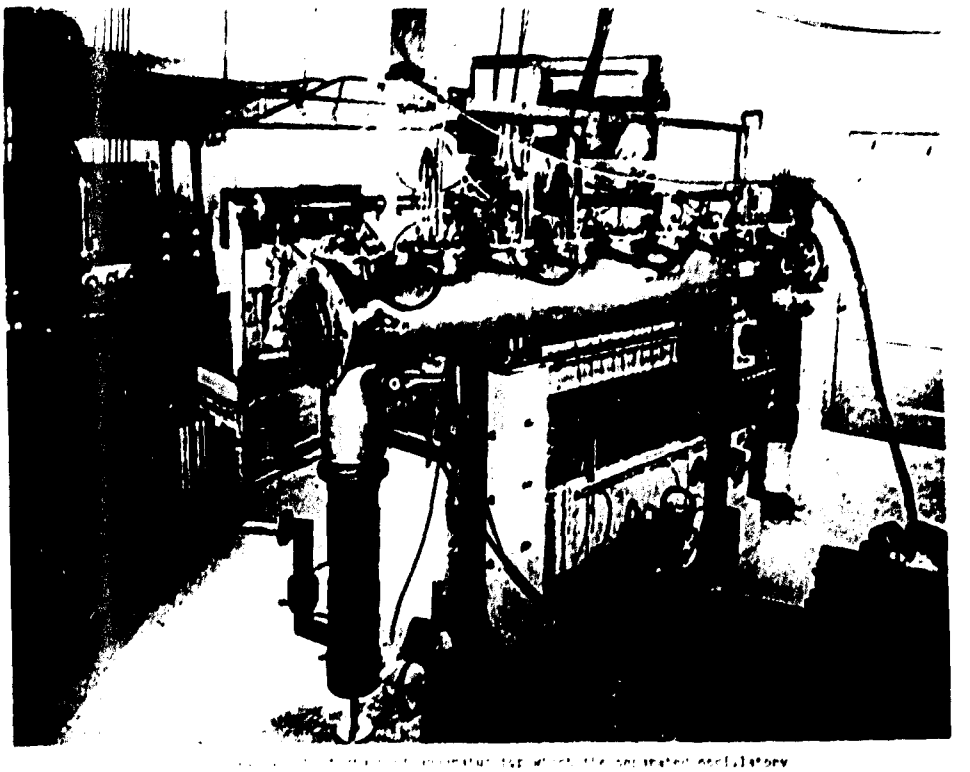


Fig. 2. Theoretical shape for separated oscillatory field resonance pattern.[18]



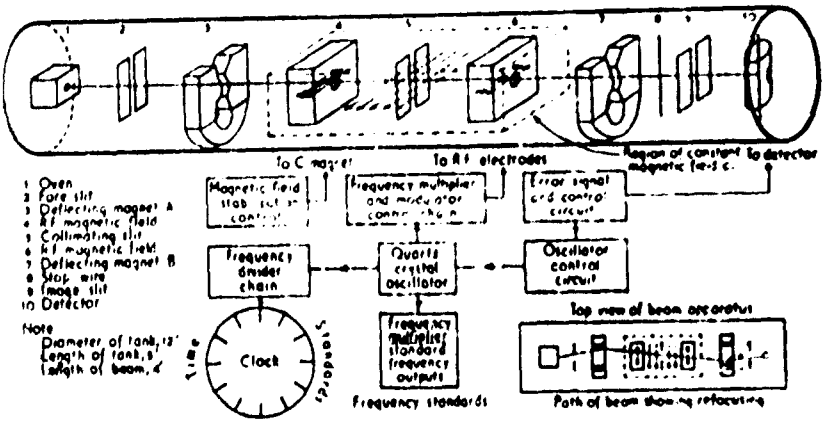


Fig. 4. Schematic diagram of a proposed atomic beam clock.[22]

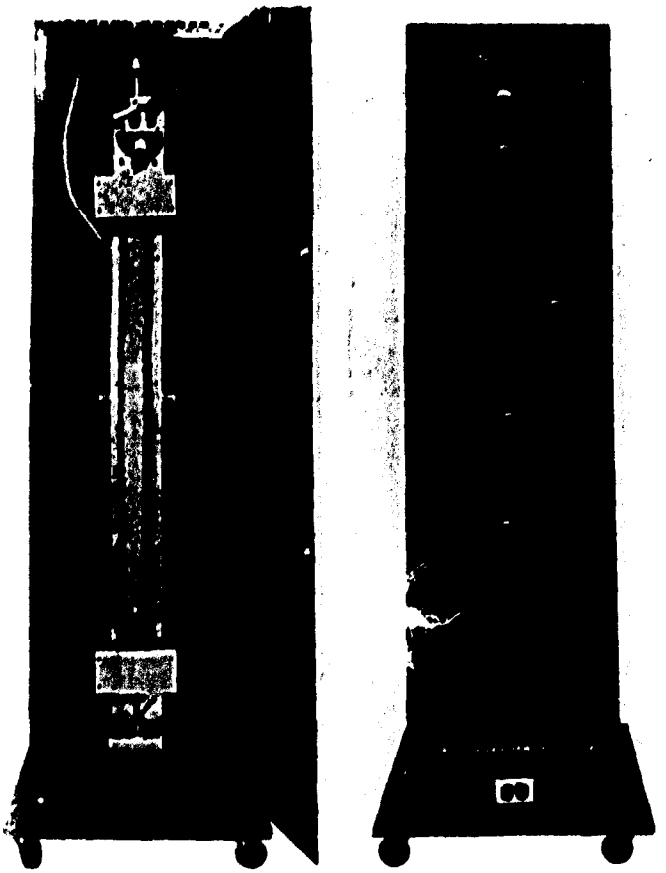


Fig. 5. Photograph of first commercial atomic beam frequency standard, the National Company Model 1000.

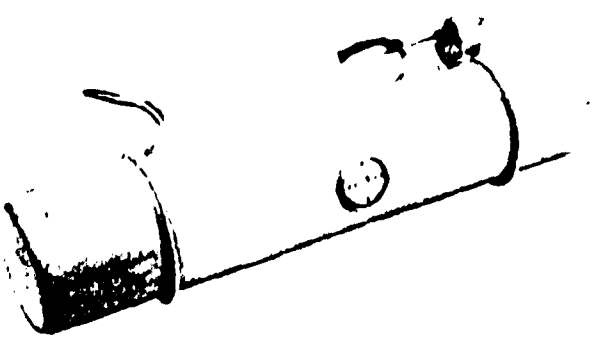


Fig. 6. Beam tube for an atomic cesium standard manufactured by Varian Associates for Hewlett Packard.

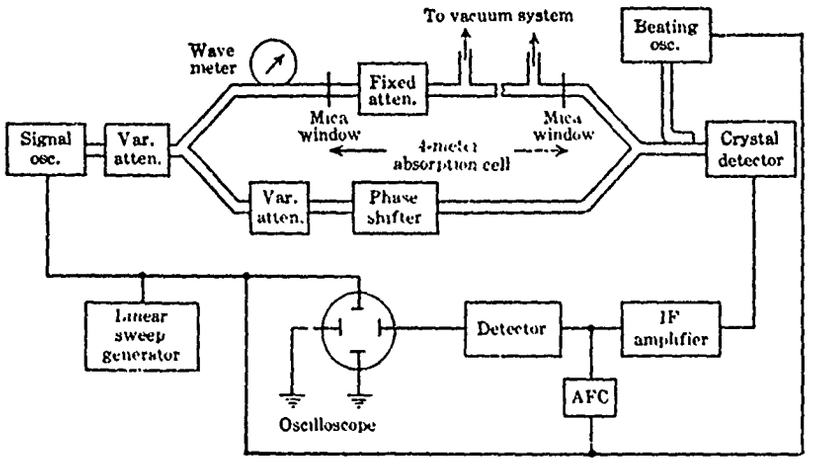
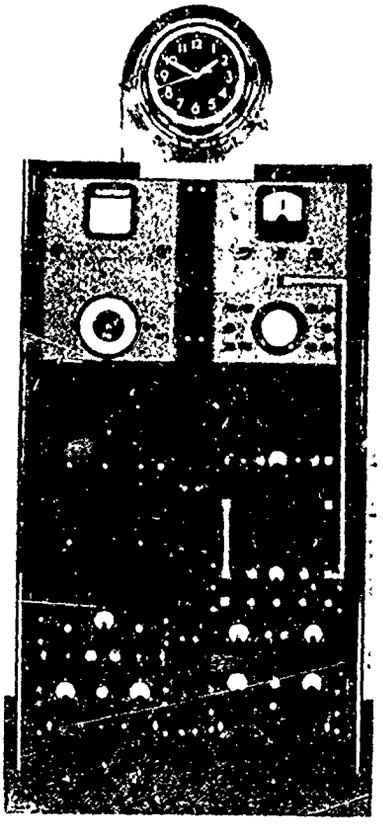
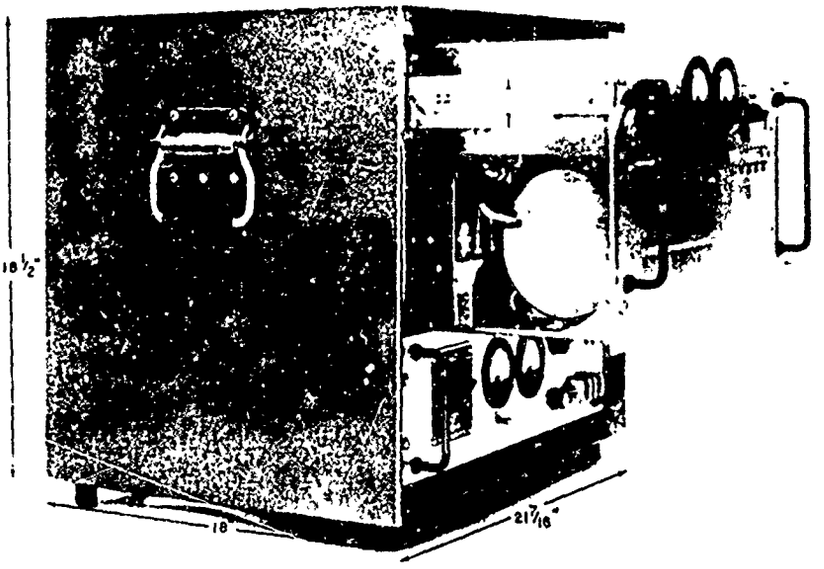


Fig. 7. A typical microwave absorption experiment using a radio-frequency bridge and heterodyne detection.





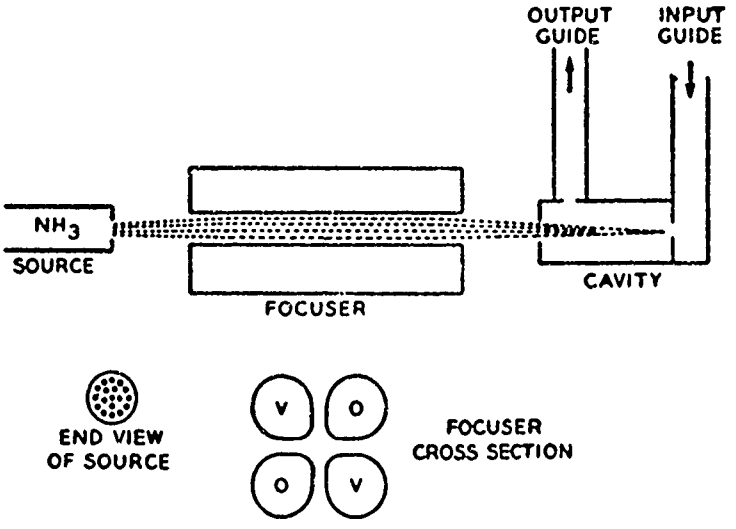
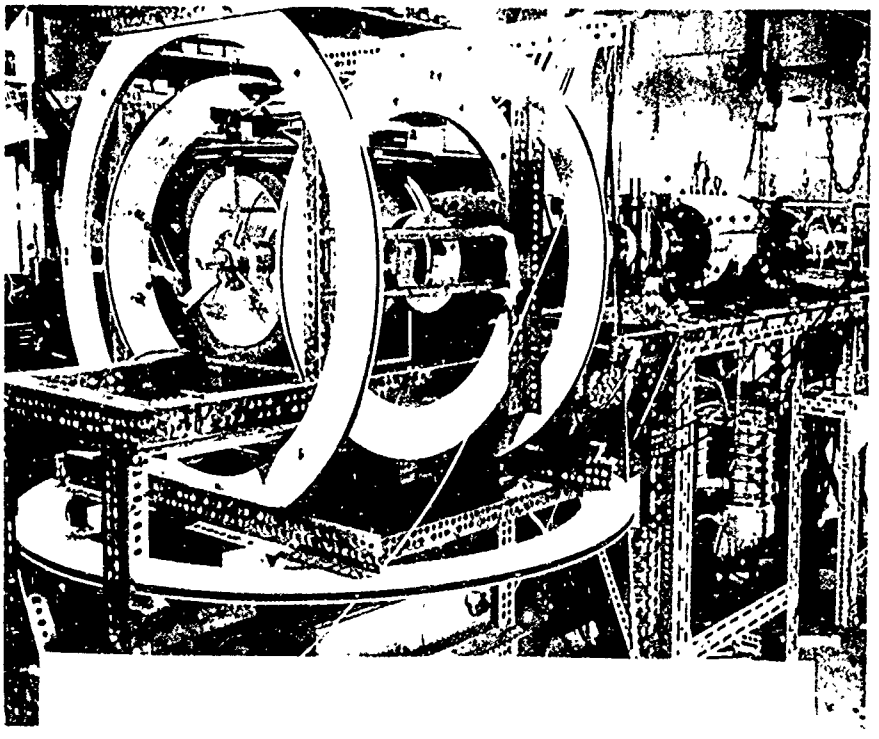


Fig. 10. Schematic diagram of original ammonia maser.[37]



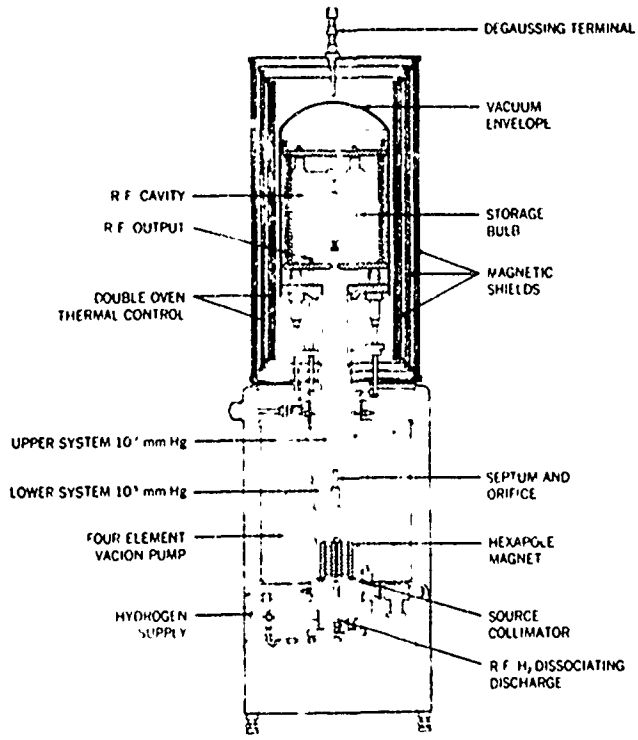
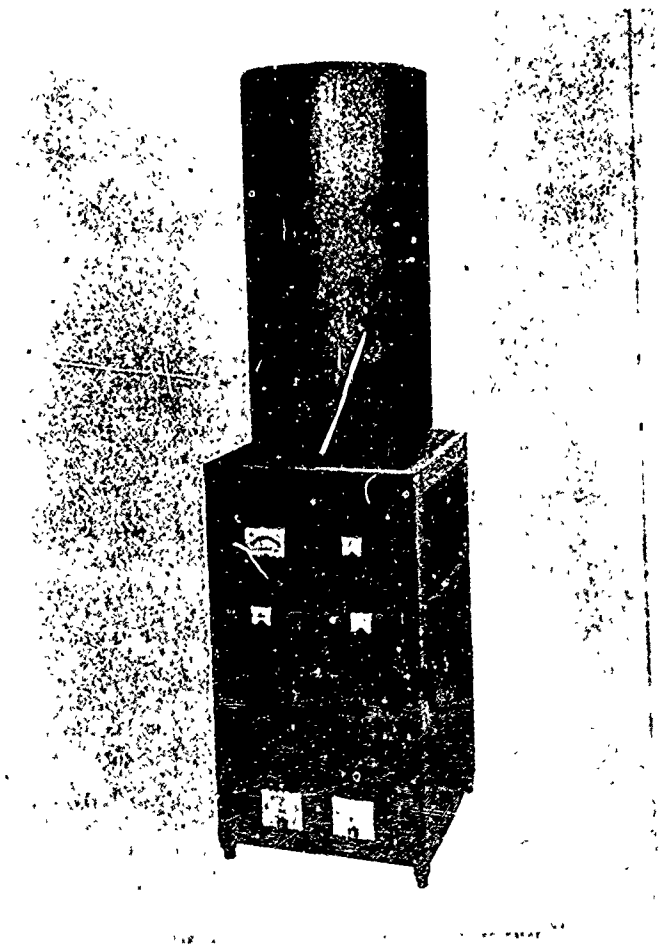


Fig. 12. Schematic diagram of commercial hydrogen maser developed by Vessot and his associates.[49]



ACOUSTICAL AND OPTICAL ACTIVITIY IN ALPHA QUARTZ

R. D. Mindlin
Department of Civil Engineering
Columbia University
New York, New York

and

R. A. Toupin
IBM Thomas J. Watson Research Center
Yorktown Heights, New York

Abstract

It is shown that a theory of elastic dielectrics, in which the stored electromechanical energy depends on the polarization gradient, accounts for both acoustical and optical activity. Formulas for the acoustical and optical rotatory powers of α-quartz are derived and the values of new material constants appearing in them are calculated from experimental data.

Introduction

Acoustical activity (rotation of the direction of mechanical displacement along the path of a transverse, elastic wave) has recently been observed by Pine [1] and by Joffrin and Levelut [2] in α-quartz. The possibility of the phenomenon appears first to have been mentioned by Silin [3]. It was accounted for by one of us [4] on the basis of the theory of elasticity in which the stored energy is a function of the gradient of the strain, in addition to the strain. Portigal and Burstein [5] found an equivalent result by assigning dependence of the elastic stiffness on the wave vector. The purpose of the present paper is to show that both acoustical and optical activity are accounted for in the theory of elastic dielectrics in which the stored electromechanical energy is a function of the polarization gradient [6] in addition to the usual strain and polarization. It has already been shown [7] that the differential equations of the resulting theory, rather than the equations of the classical theory of piezoelectricity, are the correct, long wave, low frequency limit of the finite difference equations of a lattice of shell model atoms if the shell-shell interaction between adjacent atoms is taken into account. It has also been shown that the polarization gradient can account [6] for the surface energy of deformation and polarization and it can also account [7] for an anomaly observed in measurements of electrical capacitance of thin, dielectric films. In the present paper, the field equations are exhibited for the coupled elastic-electric-magnetic system, the problem of shear waves along the trigonal axis of α-quartz is solved, formulas are obtained for the optical and acoustical rotatory powers, and numerical values of the new material constants, in the formulas, are calculated from experimental data. Essentially, the theory has it that the appearance of optical activity depends on an interaction between the polarization and the polarization gradient; the appearance of acoustical activity depends on interactions of the strain with both the polarization and the polarization gradient, and is absent if either interaction is missing – provided, of course, that the phenomenon does not depend on the strain gradient, as assumed, tentatively, in this paper.

Coupled Elastic, Electric and Magnetic Fields

The linear equations of an elastic, dielectric continuum, with the contribution of the polarization gradient taken into account, but without the coupling to the magnetic field, may be written as [6]

$$T_{ij,i} = \rho \ddot{u}_j,$$

$$E_{ij,i} + E^L_j + E_j = 0,$$

$$\epsilon_{ijk} E_{k,j} = 0 \qquad (1)$$

$$\epsilon_o E_{i,i} + P_{i,i} = 0,$$

where

$$T_{ij} = \partial W^L/\partial S_{ij},\ E^L_j = -\partial W^L/\partial P_j,\ E_{ij} = \partial W^L/\partial P_{j,i} \qquad (2)$$

and

$$W^L = \tfrac{1}{2}a_{ij}P_iP_j + \tfrac{1}{2}b_{ijk\ell}P_{j,i}P_{\ell,k} + \tfrac{1}{2}c^P_{ijk\ell}S_{ij}S_{k\ell} + d_{ijk\ell}P_{j,i}S_{k\ell} + f_{ijk}P_iS_{jk} + j_{ijk}P_iP_{k,j}, \qquad (3)$$

$$S_{ij} = \tfrac{1}{2}(u_{j,i} + u_{i,j}) \qquad (4)$$

In (1), u_i is the mechanical displacement, P_i is the polarization density, E_i is the Maxwell electric self-field, ρ is the mass density, ϵ_o is the permittivity of a vacuum and ϵ_{ijk} is the alternating tensor. In (3), W^L is the stored energy density of deformation and polarization in which $b_{ijk\ell}$, $d_{ijk\ell}$ and j_{ijk} are constants associated with the polarization gradient, $P_{j,i}$, while a_{ij}, f_{ijk} and $c^P_{ijk\ell}$ belong to the classical theory of piezoelectricity and are related to the I.R.E. Standard [8] symbols for the reciprocal dielectric susceptibility, χ_{ij}, the piezoelectric stress constant, e_{ijk}, and the elastic stiffness at constant electric field, $c^E_{ijk\ell}$, according to [9]

$$a_{ij} = \epsilon_o^{-1}\chi_{ij},\ f_{ijk} = -\epsilon_o^{-1}\chi_{i\ell}\,e_{\ell jk},\ c^P_{ijk\ell} = c^E_{ijk\ell} + \epsilon_o^{-1}\chi_{mn}e_{mij}e_{nk\ell}. \qquad (5)$$

To couple (1) to the equations of the magnetic field, it is only necessary to replace the third of (1) by

$$\epsilon_{ijk}E_{k,j} + \dot{B}_i = 0 \qquad (6)$$

and add the equations

$$\mu_o^{-1} \epsilon_{ijk} B_{k,j} - \epsilon_o \dot{E}_i - \dot{P}_i = 0, \qquad (7)$$

$$B_{i,i} = 0, \qquad (8)$$

where B_i is the magnetic flux density and μ_o is the magnetic permeability, assumed to be that of a vacuum.

It is convenient to eliminate B_i by subtracting the curl of (6) from the time derivative of (7), with the result

$$E_{j,ii} - E_{i,ij} = \epsilon_o \mu_o \ddot{E}_j + \mu_o \ddot{P}_j. \qquad (9)$$

The last of (1) and (8) are not independent of (7) and (6), respectively, and may be disregarded for the present purpose.

Thus, (9), along with the first two of (1):

$$T_{ij,i} = \rho \ddot{u}_j,$$

$$E_{ij,i} + E_j^L + E_j = 0, \qquad (10)$$

$$E_{j,ii} - E_{i,ji} = \epsilon_o \mu_o \ddot{E}_j + \mu_o \ddot{P}_j,$$

are the field equations governing mechanical and electromagnetic waves, coupled through the constitutive equations:

$$T_{ij} = c^P_{ijkl} S_{kl} + f_{kij} P_k + d_{klij} P_{l,k},$$

$$-E_j^L = f_{jkl} S_{kl} + a_{jk} P_k + j_{jkl} P_{l,k}, \qquad (11)$$

$$E_{ij} = d_{ijkl} S_{kl} + j_{kij} P_k + b_{ijkl} P_{l,k},$$

which are obtained from (2) and (3).

For a crystal of Class 32 (International) or D_3 (Schoenflies) [10], to which α-quartz, belongs, the constitutive equations take the form shown in Fig. 1, in which x_3 has been taken as the trigonal axis and x_1 one of the digonal axes. In Fig. 1, an abridged notation is used for the subscripts attached to the material constants:

11 —1, 22 —2, 33 —3, 23 —4, 32 —7, 31 —5, 13 —8, 12 —6, 21 —9.

Plane Waves Along the Trigonal Axis

We consider shear waves propagating along the x_3-axis; i.e., u_1, P_1 and E_1 are functions of x_3 and t only, but u_3, P_3, and E_3 are zero. Then the field equations (10) reduce to

$$T_{13,3} = \rho \ddot{u}_1, \qquad T_{23,3} = \rho \ddot{u}_2, \qquad (12)$$

$$E_{31,3} + E_1^L + E_1 = 0, \qquad E_{32,3} + E_2^L + E_2 = 0,$$

$$E_{1,33} = \epsilon_o \mu_o \ddot{E}_1 + \mu_o \ddot{P}_1, \qquad E_{2,33} = \epsilon_o \mu_o \ddot{E}_2 + \mu_o \ddot{P}_2$$

and the constitutive equations (11) reduce to

$$T_{32} = c^P_{44} u_{2,3} + f_{14} P_1 + d_{74} P_{2,3},$$

$$T_{31} = c^P_{44} u_{1,3} - f_{14} P_2 + d_{74} P_{1,3},$$

$$-E_1^L = f_{14} u_{2,3} + \epsilon_o^{-1} \chi_{11} P_1 + j_{17} P_{2,3},$$

$$-E_2^L = -f_{14} u_{1,3} + \epsilon_o^{-1} \chi_{11} P_2 - j_{17} P_{1,3}, \qquad (13)$$

$$E_{32} = d_{74} u_{2,3} + j_{17} P_1 + b_{55} P_{2,3},$$

$$E_{31} = d_{74} u_{1,3} - j_{17} P_2 + b_{55} P_{1,3}.$$

Inserting (13) in (12), we have

$$c^P_{44} u_{1,33} - f_{14} P_{2,3} + d_{74} P_{1,33} = \rho \ddot{u}_1,$$

$$c^P_{44} u_{2,33} + f_{14} P_{1,3} + d_{74} P_{2,33} = \rho \ddot{u}_2,$$

$$d_{74} u_{1,33} - 2j_{17} P_{2,3} + b_{55} P_{1,33} - f_{14} u_{2,3} - \epsilon_o^{-1} \chi_{11} P_1 + E_1 = 0, \qquad (14)$$

$$d_{74} u_{2,33} + 2j_{17} P_{1,3} + b_{55} P_{2,33} + f_{14} u_{1,3} - \epsilon_o^{-1} \chi_{11} P_2 + E_2 = 0,$$

$$E_{1,33} = \epsilon_o \mu_o \ddot{E}_1 + \mu_o \ddot{P}_1,$$

$$E_{2,33} = \epsilon_o \mu_o \ddot{E}_2 + \mu_o \ddot{P}_2.$$

Now, take (15)

$$u_1 = A_1 \sin \zeta (x_3 - vt), \qquad u_2 = A_2 \cos \zeta (x_3 - vt),$$

$$P_1 = B_1 \sin \zeta (x_3 - vt), \qquad P_2 = B_2 \cos \zeta (x_3 - vt),$$

$$E_1 = C_1 \sin \zeta (x_3 - vt), \qquad E_2 = C_2 \cos \zeta (x_3 - vt)$$

and substitute these functions in (14) to find

$$(c^P_{44} - \rho v^2) \zeta A_1 - f_{14} B_2 + d_{74} \zeta B_1 = 0,$$

$$(c^P_{44} - \rho v^2) \zeta A_2 - f_{14} B_1 + d_{74} \zeta B_2 = 0, \qquad (16)$$

$$d_{74} \zeta^2 A_1 - f_{14} \zeta A_2 + (b_{55} \zeta^2 + \epsilon_o^{-1} \chi_{11}) B_1 - 2j_{17} \zeta B_2 - C_1 = 0,$$

$$d_{74} \zeta^2 A_2 - f_{14} \zeta A_1 + (b_{55} \zeta^2 + \epsilon_o^{-1} \chi_{11}) B_2 - 2j_{17} \zeta B_1 - C_2 = 0,$$

$$\mu_o v^2 B_1 + (\epsilon_o \mu_o v^2 - 1) C_1 = 0,$$

$$\mu_o v^2 B_2 + (\epsilon_o \mu_o v^2 - 1) C_2 = 0.$$

Adding and subtracting these equations in pairs, we have

$$(c_{44}^P - \rho v^2)(A_1 \pm A_2) + (d_{74}\zeta \mp f_{14})(B_1 \pm B_2) = 0,$$

$$(d_{74}\zeta \mp f_{14})(A_1 \pm A_2) \qquad (17)$$

$$+ (b_{55}\zeta^2 + \epsilon_o^{-1}\chi_{11} \mp 2j_{17}\zeta)(B_1 \pm B_2) - (C_1 \pm C_2) = 0,$$

$$\mu_o v^2 (B_1 \pm B_2) + (\epsilon_o \mu_o v^2 - 1)(C_1 \pm C_2) = 0.$$

Thus, there are two solutions, each corresponding to circularly polarized waves [11, p.222] since the amplitudes must satisfy

$$A_1 = \pm A_2, \quad B_1 = \pm B_2, \quad C_1 = \pm C_2 \qquad (18)$$

with either all upper signs or all lower signs. Upon substituting (18) into (15), we see that the upper and lower signs give right and left circular polarization, respectively. The velocities are obtained by setting the determinant of the coefficients of the amplitudes in (17) equal to zero:

$$\begin{vmatrix} c_{44}^P - \rho v^2 & d_{74}\zeta \mp f_{14} & 0 \\ d_{74}\zeta \mp f_{14} & b_{55}\zeta^2 + \epsilon_o^{-1}\chi_{11} \mp 2j_{17}\zeta & 1 \\ 0 & 1 & \mu_o^{-1} v^{-2} - \epsilon_o \end{vmatrix} = 0. \qquad (19)$$

This is a quadratic equation in v^2, so that there are two pairs of oppositely circularly polarized waves. Each pair of such waves combines to produce a linearly polarized wave with a rotating direction of polarization [11, p.222]. Thus, we have two cases of rotary polarization. These may be identified as optical and acoustical by separating out first the electromagnetic part of (19) and then the electromechanical part. The two may, in fact, be considered separately owing to the large ratio of the frequencies (of the order of 10^5) at which the two effects are observed.

Optical Activity

The electromagnetic part of the determinant in (19) is the minor of the upper left element. Thus, the pair of optical velocities is given by

$$\begin{vmatrix} b_{55}\zeta^2 + \epsilon_o^{-1}\chi_{11} \mp 2j_{17} & 1 \\ 1 & \mu^{-1} v^{-2} - \epsilon_o \end{vmatrix} = 0, \qquad (20)$$

which yields the dispersion formula (compare [11], p. 426)

$$n_\pm^2 - 1 = (\chi_{11} \mp 2\epsilon_o j_{17}\zeta + \epsilon_o b_{55}\zeta^2)^{-1}, \qquad (21)$$

where $n_\pm$ are the indices of refraction:

$$n_\pm = c/v_\pm \qquad (22)$$

and c is the velocity of light in vacuo:

$$c = (\epsilon_o \mu_o)^{-1}. \qquad (23)$$

From (21), we have

$$(n_-^2 - 1)^{-1} - (n_+^2 - 1)^{-1} = 4\epsilon_o j_{17}\zeta. \qquad (24)$$

Now, define $n = \frac{1}{2}(n_+ + n_-)$ and assume

$$|n_+ - n_-| \ll n_+ + n_-. \qquad (25)$$

Then (24) becomes, approximately,

$$n_+ - n_- = 2(n^2 - 1)^2 \epsilon_o j_{17}\zeta_o, \qquad (26)$$

where $\zeta_o (= \zeta/n)$ is the wave number in vacuo.

The optical rotary power, in radians per unit length, is given by [11, p. 222]

$$\theta_{OP} = \tfrac{1}{2}\zeta_o(n_- - n_+). \qquad (27)$$

Accordingly, from (26) and (27),

$$\theta_{OP} = -(n^2 - 1)^2 \epsilon_o j_{17}\zeta_o^2 \qquad (28)$$

is the formula for the optical rotatory power in terms of the average index of refraction, n, along the optic axis, the wave length in vacuo, $\lambda_o (= 2\pi/\zeta_o)$, the fundamental constant ϵ_o and the material constant $j_{17} (= j_{132})$ which, as may be seen in (3), measures the interaction between the polarization and the polarization gradient.

Acoustical Activity

The electromechanical part of the determinant in (19) is the minor of the lower right element, so that we have

$$\begin{vmatrix} c_{44}^P - \rho v^2 & d_{74}\zeta \mp f_{14} \\ d_{74}\zeta \mp f_{14} & b_{55}\zeta^2 + \epsilon_o^{-1}\chi_{11} \mp 2j_{17}\zeta \end{vmatrix} = 0 \qquad (29)$$

for the equation determining the velocities of the two acoustical waves, as influenced by the quasi-static polarization and polarization gradient.
From (29),

$$\rho v_\pm^2 = c_{44}^P - (d_{74}\zeta \mp f_{14})^2 / (b_{55}\zeta^2 + \epsilon_o^{-1}\chi_{11} \mp 2j_{17}\zeta) \qquad (30)$$

In view of (18) and the inequality of v_+ and v_-, the superposition of the two waves results in rotary polarization (acoustical activity) with acoustical rotatory power

$$\theta_{AC} = \tfrac{1}{2}\omega(v_-^{-1} - v_+^{-1}), \tag{31}$$

where ω is the circular frequency.

Both waves are dispersive. At the zero frequency (long wave) limit, from (30) and (5),

$$\lim_{\zeta\to 0} \rho v_\pm^2 = c_{44}^P - \epsilon_o f_{14}^2/\chi_{11} = c_{44}^E, \tag{32}$$

which is the result (without acoustical activity, since $v_+ = v_-$) that would be obtained if the contribution of the polarization gradient were omitted, i.e., if d_{74}, b_{55} and j_{17} were assumed to be zero. As the frequency increases from zero, the absolute velocity difference, $|v_+ - v_-|$, at first increases, so that the acoustical activity appears and increases. With further increases of frequency, the velocity difference again approaches zero, since

$$\lim_{\zeta\to\infty} \rho v_\pm^2 = c_{44}^P - d_{74}^2/b_{55}, \tag{33}$$

so that the acoustical activity diminishes; but this is undoubtedly beyond the range of applicability of the continuum theory. Up to moderately large wave numbers, (30) and (5) give, to the first order in ζ,

$$v_0/v_\pm = 1 \mp (d_{74} - j_{17}e_{14})e_{14}\zeta/c_{44}^E \tag{34}$$

where $v_0^2 = c_{44}^E/\rho$. In this range, the frequency is approximately proportional to the wave number: $\omega = v_0\zeta$, so that, from (31) and (34),

$$\theta_{AC} = (d_{74} - j_{17}e_{14})e_{14}\rho\omega^2/(c_{44}^E)^2. \tag{35}$$

Thus, at frequencies up to, say, 10^{10} cps, the acoustical rotatory power is approximately proportional to the square of the frequency and depends on the constants ρ, e_{14} and c_{44}^E, which are commonly encountered in piezoelectricity theory, and also on the constants d_{74} and j_{17} which control the coupling of the polarization gradient with the strain and polarization, respectively.

Application to Quartz

For α-quartz, all the quantities in the formula (28) for optical rotatory power are known except j_{17}. Thus, for left-handed quartz and the sodium D line,

θ_{0P} = -379 radian/meter [11, p. 481],

λ_o = 5893×10^{-10} meter [11, p. 481],

n = 1.5533 [11, p. 481],

ϵ_o = 8.854×10^{-12} farad/meter [10].

Hence,

$$j_{17} = -\theta_{0P}\lambda_o^2/4\pi^2\epsilon_o(n^2-1)^2 = 0.19 \text{ meter}^2/\text{farad}. \tag{36}$$

With the value of j_{17} known, all quantities in the formula (35) for acoustical rotatory power are known except d_{74}. In particular, Pine [1] finds that the acoustical and optical activities have opposite signs and the acoustical rotary power along the trigonal axis is about 220 radians/meter at one gigahertz. Thus, for left handed α-quartz,

θ_{AC} = 220 radian/meter [1],

ω = $2\pi \times 10^9$ radian/second [1],

ρ = 2.65×10^3 kilogram/meter^3 [12],

c_{44}^E = 57.94×10^9 newton/meter^2 [13],

e_{14} = -0.0406 coulomb/meter^2 [13].

Hence,

$$d_{74} = \theta_{AC}(c_{44}^E)^2/e_{14}\rho\omega^2 + j_{17}e_{14} = -174 - 0.0077 \text{ volt} \tag{37}$$

The second term, $j_{17}e_{14}$, is negligible in comparison with the first, so that we may drop the dependence of acoustical rotatory power on j_{17} and replace (35) with

$$\theta_{AC} = d_{74}e_{14}\rho\omega^2/(c_{44}^E)^2. \tag{39}$$

Thus, according to this theory, the presence of acoustical activity in α-quartz depends on the existence of the piezoelectric stress constant $e_{14}(=e_{123})$ and the interaction constant $d_{74}(=d_{3223})$ between strain and polarization gradient, whereas the presence of optical activity depends only on the existence of the interaction constant j_{17} $(=j_{132})$ between polarization and polarization gradient.

References

[1] A. S. Pine, "Direct Observation of Acoustical Activity in α-Quartz", Phys. Rev. B, 2, 2049-2054 (1970).

[2] J. Joffrin and A. Levelut, "Mise en évidence et mesure du pouvoir rotatoire acoustique naturel du quartz-α", Laboratoire d'Ultrasons, Faculté des Sciences de Paris, forthcoming.

[3] V. P. Silin, 'Contribution to the theory of absorption of ultrasound in metals", JETP (U. S. S. R.) 38, 977-983 (1960); 11, 703-707, (1960) of Am. Phys. Soc. Trans. Soviet Physics JETP.

[4] R. A. Toupin, "Elastic Materials with Couple-Stresses", Arch. Rat. Mech. Anal., 11, 385-414 (1962).

[5] D. L. Portigal and E. Burstein, 'Acoustical Activity and Other First-Order Spatial Dispersion Effects in Crystals:, Phys. Rev. 10, 673-678 (1968).

[6] R. D. Mindlin, "Polarization Gradient in Elastic Dielectrics", Int. J. Solids Structures, 4, 637-642 (1968).

[7] R. D. Mindlin, "Continuum and Lattice Theories of Influence of Electromechanical Coupling on Capacitance of Thin, Dielectric Films", Int. J. Solids Structures, 5, 1197-1208 (1969).

[8] Standards on Piezoelectric Crystals, 1949, Proc. Inst. Radio Engrs. 87, 1378-1385 (1949).

[9] R. D. Mindlin, 'Lectures on Polarization Gradient in Elastic Dielectrics," Int. Centre for Mechanical Sciences, Udine, Italy, 1970.

[10] J. F. Nye, Physical Properties of Crystals, Oxford, 1959.

[11] R. A. Houston, A Treatise on Light, Longmans, Green and Co., London, 1934.

[12] W. G. Cady, Piezoelectricity, McGraw-Hill Book Co., New York, 1946.

[13] R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz", Phys. Rev. 110, 1060-1061 (1958.

FIGURE 1. Constitutive relations for crystal class 32 (D_3).

	S_{11}	S_{22}	S_{33}	$2S_{23}$	$2S_{31}$	$2S_{12}$	P_1	P_2	P_3	$P_{1,1}$	$P_{2,2}$	$P_{3,3}$	$P_{3,2}$	$P_{2,3}$	$P_{1,3}$	$P_{3,1}$	$P_{2,1}$	$P_{1,2}$
T_{11}	c^P_{11}	c^P_{12}	c^P_{13}	c^P_{14}	0	0	f_{11}	0	0	d_{11}	d_{12}	d_{31}	d_{41}	d_{71}	0	0	0	0
T_{22}	c^P_{12}	c^P_{11}	c^P_{13}	$-c^P_{14}$	0	0	$-f_{11}$	0	0	d_{12}	d_{11}	d_{31}	$-d_{41}$	$-d_{71}$	0	0	0	0
T_{33}	c^P_{13}	c^P_{13}	c^P_{33}	0	0	0	0	0	0	d_{13}	d_{13}	d_{33}	0	0	0	0	0	0
T_{23}	c^P_{14}	$-c^P_{14}$	0	c^P_{44}	0	0	f_{14}	0	0	d_{14}	$-d_{14}$	0	d_{44}	d_{74}	0	0	0	0
T_{31}	0	0	0	0	c^P_{44}	c^P_{14}	0	$-f_{14}$	0	0	0	0	0	0	d_{74}	d_{44}	d_{14}	d_{14}
T_{12}	0	0	0	0	c^P_{14}	c^P_{66}	0	$-f_{11}$	0	0	0	0	0	0	d_{71}	d_{41}	d_{66}	d_{66}
$-E^L_1$	f_{11}	$-f_{11}$	0	f_{14}	0	0	a_{11}	0	0	j_{11}	$-j_{11}$	0	j_{14}	j_{17}	0	0	0	0
$-E^L_2$	0	0	0	0	$-f_{14}$	$-f_{11}$	0	a_{11}	0	0	0	0	0	0	$-j_{17}$	$-j_{14}$	$-j_{11}$	$-j_{11}$
$-E^L_3$	0	0	0	0	0	0	0	0	a_{33}	0	0	0	0	0	0	0	j_{36}	$-j_{36}$
E_{11}	d_{11}	d_{12}	d_{13}	d_{14}	0	0	j_{11}	0	0	b_{11}	b_{12}	b_{13}	b_{14}	b_{17}	0	0	[illegible]	
E_{22}	d_{12}	d_{11}	d_{13}	$-d_{14}$	0	0	$-j_{11}$	0	0	b_{12}	b_{11}	b_{13}	$-b_{14}$	$-b_{17}$	0	0	[illegible]	0
E_{33}	d_{31}	d_{31}	d_{33}	0	0	0	0	0	0	b_{13}	b_{13}	b_{33}	0	[illegible]	0	0	0	0
E_{23}	d_{41}	$-d_{41}$	0	d_{44}	0	0	j_{14}	0	0	b_{14}	$-b_{14}$	[illegible]	[illegible]	b_{47}	0	0	0	0
E_{32}	d_{71}	$-d_{71}$	0	d_{74}	0	0	j_{17}	0	0	b_{17}	$-b_{17}$	[illegible]	[illegible]	b_{55}	0	0	0	0
E_{31}	0	0	0	0	d_{74}	d_{71}	0	$-j_{17}$	0	0	0	0	0	0	b_{55}	b_{47}	b_{17}	b_{17}
E_{13}	0	0	0	0	d_{44}	d_{41}	0	$-j_{14}$	0	0	0	0	0	0	b_{47}	b_{44}	b_{14}	b_{14}
E_{12}	0	0	0	0	d_{14}	d_{66}	0	$-j_{11}$	j_{36}	0	0	0	0	0	b_{17}	b_{14}	b_{66}	b_{69}
E_{21}	0	0	0	0	d_{14}	d_{66}	0	$-j_{11}$	$-j_{36}$	0	0	0	0	0	b_{17}	b_{14}	b_{69}	b_{66}

$$c^P_{66} = \tfrac{1}{2}(c^P_{11} - c^P_{12}), \qquad d_{66} = \tfrac{1}{2}(d_{11} - d_{12}), \qquad b_{66} + b_{69} = b_{11} - b_{12}$$

EXTENSIONAL, FLEXURAL AND WIDTH-SHEAR VIBRATIONS OF THIN RECTANGULAR CRYSTAL PLATES

P. C. Y. Lee
Princeton University
Princeton, New Jersey 08540

Summary

An approximate, one-dimensional theory is derived for thin crystal plates or bars with narrow rectangular cross section, and for cases in which the stiffness matrix of the crystal c_{ij} referring to the plate axes exhibits monoclinic symmetry. A set of five equations of motions is separated into two groups: the extensional, width-length flexural, and width-shear motion and the transverse shear, and width-twist motion.

Closed form solutions of coupled extensional, flexural, and width-shear vibration are obtained for thin rectangular plate with free edges. Calculated resonant frequencies as a function of length-to-width ratio of the plate are compared with detailed measurements by Miss D.M. Spears on +5° X-cut, quartz plates. For the use in design, explicit algebraic formulas for predicting extensional and flexural frequencies are obtained in terms of elastic stiffnesses, density, and width-to-length ratio of the plates.

Introduction

Thin rectangular plates or finite bars with narrow, rectangular cross section (Fig. 1) excited mainly in extensional and flexural modes of vibration have been used extensively as piezoelectric resonators in filter circuit applications. For very low frequencies ($\Omega \ll 1$) and small width-to-length ratio ($b/c \ll 1$) of the plate, the classical extensional equation and flexural equation including the effects of rotary and lateral inertia[1] have been shown to predict resonant frequencies with great accuracy. However, due to the anisotropy of the crystals, extensional and flexural modes are sometimes strongly coupled, as occurs in the case of Φ X-cut, quartz plates. Furthermore, in case of plates being excited in higher flexural modes or with larger values of width-to-length ratio, the effect due to the presence of the width-shear mode becomes of more importance. Coupled longitudinal and flexural vibrations of both thin rectangular and parallelogram plates have been studied extensively by Jumonji and Onoe.[2] In their analysis, approximate solutions are obtained by linear combination of Lamb waves or solutions corresponding to frequency branches of the Rayleigh-Lamb dispersion equation. Hence the accuracy of their approximation can be improved by taking additional higher branches into account.

The purpose of this paper is to formulate an approximate theory which includes extensional, flexural and width-shear modes and allows closed form solutions to satisfy free edge conditions. First, a set of approximate, one-dimensional plate equations for generalized plane stress is derived based on Mindlin's first-order, two-dimensional equations for crystal plates.[3] When the stiffness matrix C_{ij} of the crystal referring to plate axes exhibits monoclinic symmetry, such as in rotated Y-cut, or Φ °X-cut quartz plates, this set of five one-dimensional equations is separated into two groups corresponding to extension, width-length flexure and width-shear motion, and transverse shear and width-twist motion. In case the crystal is plated with electrode platings, the mechanical effect of the platings has been taken into account.

Then closed form solutions are obtained for thin rectangular plates with free edges, which are vibrating in extensional, flexural, and width-shear modes. The resonant frequencies are computed as a function of length-to-width ratio while the width of the plate and the electrode mass are kept at constant values. The predicted result is compared to D.M. Spears' detailed data on +5° X-cut, quartz plates.[4] For use in design, explicit algebraic formulas for predicting extensional and flexural frequencies separately are obtained in terms of elastic stiffness, density, and the length-to-width ratio of the plates.

One-Dimensional Plate Equations

Consider a rectangular plate of thickness 2a, width 2b, and length 2c in the rectangular coordinate system (x_i) as shown in Fig. 1. Mindlin's two-dimensional, first order, stress equations of motion for crystal plates, which take into account the first five lowest vibrational modes, are:[3]

$$
\begin{aligned}
T^{(0)}_{11,1} + T^{(0)}_{13,3} &= 2b\rho \frac{\partial^2 u_1^{(0)}}{\partial t^2}, \\
T^{(0)}_{12,1} + T^{(0)}_{23,3} &= 2b\rho \frac{\partial^2 u_2^{(0)}}{\partial t^2}, \\
T^{(0)}_{13,1} + T^{(0)}_{33,3} &= 2b\rho \frac{\partial^2 u_3^{(0)}}{\partial t^2}, \\
T^{(1)}_{11,1} + T^{(1)}_{13,3} - T^{(0)}_{32} &= \frac{2}{3} b^3 \rho \frac{\partial^2 u_1^{(1)}}{\partial t^2}, \\
T^{(1)}_{13,1} + T^{(1)}_{33,3} - T^{(0)}_{32} &= \frac{2}{3} b^3 \rho \frac{\partial^2 u_3^{(1)}}{\partial t^2},
\end{aligned}
\tag{1}
$$

where $u_i^{(0)}$ are the components of the displacement and $u_i^{(1)}$ are the components of the rotation in the x_i direction of a plate element which is originally parallel to the x_2 axis. $T_{ij}^{(0)}$ are stress resultants and $T_{ij}^{(1)}$ are stress couples.

The stress-strain relations with monoclinic symmetry, such as the case for rotated Y-cut or Φ ° X-cut of quartz are:

$$T_{11}^{(0)} = 2b[\bar{c}_{11}u_{1,1}^{(0)} + \bar{c}_{13}u_{3,3}^{(0)} + k_3\bar{c}_{14}(u_{2,3}^{(0)} + u_3^{(1)})],$$

$$T_{33}^{(0)} = 2b[\bar{c}_{13}u_{1,1}^{(0)} + \bar{c}_{33}u_{3,3}^{(0)} + k_3\bar{c}_{34}(u_{2,3}^{(0)} + u_3^{(1)})],$$

$$T_{13}^{(0)} = 2b[c_{55}(u_{3,1}^{(0)} + u_{1,3}^{(0)} + k_1c_{56}(u_{2,1}^{(0)} + u_1^{(1)})], \qquad (2)$$

$$T_{12}^{(0)} = 2b[k_1c_{56}(u_{3,1}^{(0)} + u_{1,3}^{(0)}) + k_1^2\,c_{66}(u_{2,1}^{(0)} + u_1^{(1)})],$$

$$T_{32}^{(0)} = 2b[k_3\bar{c}_{14}\,u_{1,1}^{(0)} + k_3\bar{c}_{34}u_{3,3}^{(0)} + k_3^2\,\bar{c}_{44}(u_{2,3}^{(0)} + u_3^{(1)})],$$

$$T_{11}^{(1)} = \frac{2}{3}b^3(\gamma_{11}u_{1,1}^{(1)} + \gamma_{13}u_{3,3}^{(1)}),$$

$$T_{13}^{(1)} = \frac{2}{3}b^3\gamma_{55}(u_{3,1}^{(1)} + u_{1,3}^{(1)}),$$

where

$$k_1^2 = \pi^2/12,$$

$$k_3^2 = \pi^2\{c_{22} + c_{44} - [(c_{22} - c_{44})^2 + 4c_{24}^2]^{1/2}\}/24\bar{c}_{44},$$

$$\bar{c}_{pq} = c_{pq} - c_{2p}c_{q2}/c_{22}, \qquad p,q = 1,3,4$$

$$\gamma_{pq} = \bar{c}_{pq} - \bar{c}_{4p}\,\bar{c}_{q4}/\bar{c}_{44}, \qquad p,q = 1,3 \qquad (3)$$

$$\gamma_{55} = c_{55} - c_{56}^2/c_{66}.$$

These equations are two-dimensional and the traction free edge conditions at $x_2 = \pm b$ have been accounted for. When the thickness of the plate 2a is small as compared with wave lengths and other plate dimensions, the components of displacement and stress can be approximated by their average values across the thickness as defined respectively by[5]

$$v_j^{(n)}(x_3,t) = \frac{1}{2a}\int_{-a}^{a} u_j^{(n)}(x_1,x_3,t)\,dx_1,$$

$$\sigma_{ij}^{(n)}(x_3,t) = \frac{1}{2a}\int_{-a}^{a} T_{ij}^{(n)}(x_1,x_3,t)\,dx_1. \qquad (4)$$

Multiplying (1) by 1/2a and integrating through the thickness, one obtains:

$$\sigma_{13,3}^{(0)} + H_1^{(0)} = 2b\rho\frac{\partial^2 v_1^{(0)}}{\partial t^2},$$

$$\sigma_{23,3}^{(0)} + H_2^{(0)} = 2b\rho\frac{\partial^2 v_2^{(0)}}{\partial t^2},$$

$$\sigma_{33,3}^{(0)} + H_3^{(0)} = 2b\rho\frac{\partial^2 v_3^{(0)}}{\partial t^2}, \qquad (5)$$

$$\sigma_{13,3}^{(1)} - \sigma_{12}^{(0)} + H_1^{(1)} = \frac{2}{3}b^3\rho\frac{\partial v_1^{(1)}}{\partial t^2},$$

$$\sigma_{33,3}^{(1)} - \sigma_{32}^{(0)} + H_3^{(1)} = \frac{2}{3}b^3\rho\frac{\partial v_3^{(1)}}{\partial t^2},$$

where,

$$H_j^{(n)} = \frac{1}{2a}[T_{1j}^{(n)}(a) - T_{1j}^{(n)}(-a)]$$

are the stress resultants (for n=0) and stress couples (for n=1) on the faces of the plate at $x_1 = \pm a$. For plates with traction-free faces, the state of stress is a case of generalized plane stress. The proper conditions for traction-free face conditions are[6]

$$T_{11}^{(0)} = 0, \qquad T_{11}^{(1)} = 0, \qquad (7)$$

and

$$T_{12}^{(0)}\Big|_{x_1=\pm a} = 0, \quad T_{13}^{(0)}\Big|_{x_1=\pm a} = 0, \quad T_{13}^{(1)}\Big|_{x_1=\pm a} = 0. \qquad (8)$$

It is seen that (7) and (8) imply that $H_j^{(n)} = 0$. Furthermore for symmetric extensional and flexural motion in the x_3 direction, $u_3^{(0)}$, $u_3^{(1)}$, and $u_2^{(0)}$ are assumed to be symmetric with respect to the x_2x_3 plane; therefore

$$\int_{-a}^{a}[u_3^{(0)}, u_3^{(1)}, u_2^{(0)}]_{,1}\,dx_1 = 0, \qquad (9)$$

From (2) and (7), one obtains:

$$u_{1,1}^{(0)} = \frac{\bar{c}_{13}}{\bar{c}_{11}}u_{3,3}^{(0)} - k_3\frac{\bar{c}_{14}}{\bar{c}_{11}}(u_{2,3}^{(0)} + u_3^{(1)}),$$

$$u_{1,1}^{(1)} = -\frac{\gamma_{13}}{\gamma_{11}}u_{3,3}^{(1)} \qquad (10)$$

Substitution of (10) into (2), integration through the thickness, and utilization of the assumption (9) yields the stress-strain relations for generalized plane stress theory:

$$\sigma_{33}^{(0)} = 2b[c_{33}^{*}v_{3,3}^{(0)} + k_3c_{34}^{*}(v_{2,3}^{(0)} + v_3^{(1)})],$$

$$\sigma_{32}^{(0)} = 2b[k_3c_{34}^{*}v_{2,3}^{(0)} + k_3^2c_{44}^{*}(v_{2,3}^{(0)} + v_3^{(1)})], \qquad (11a)$$

$$\sigma_{33}^{(1)} = \frac{2}{3}b^3\gamma_{33}^{*}v_{3,3}^{(1)},$$

and

$$\sigma_{13}^{(0)} = 2b(c_{55}v_{1,3}^{(0)} + k_1c_{56}v_1^{(1)}),$$

$$\sigma_{12}^{(0)} = 2b(k_1c_{56}v_{1,3}^{(0)} + k_1^2c_{66}v_1^{(1)}), \qquad (12a)$$

$$\sigma_{13}^{(1)} = \frac{2}{3}b^3\gamma_{55}v_{1,3}^{(1)},$$

where

$$c_{pq}^{*} = \bar{c}_{pq} - \bar{c}_{1p}\,\bar{c}_{q1}/\bar{c}_{11}, \quad \gamma_{33}^{*} = \gamma_{33} - \gamma_{13}^2/\gamma_{11}, \qquad (13)$$
$$p,q = 3,4$$

The displacement equations of motion are obtained by substitution of (11a) and (12a) into (5):

$$c_{33}^{*}v_{3,33}^{(0)} + k_3c_{34}^{*}(v_{2,33}^{(0)} + v_{3,3}^{(1)}) = \rho\frac{\partial^2 v_3^{(0)}}{\partial t^2},$$

$$k_3c_{14}^{*}v_{3,33}^{(0)} + k_3^2c_{44}(v_{2,33}^{(0)} + v_{3,3}^{(1)}) = \rho\frac{\partial^2 v_2^{(0)}}{\partial t^2}, \qquad (11b)$$

$$\gamma_{33}^{*}v_{3,3}^{(1)} - 3b^{-2}[k_3c_{34}^{*}v_{3,3}^{(0)} + k_3^2c_{44}^{*}(v_{2,3}^{(0)} + v_3^{(1)})] = \frac{\partial^2 v_3^{(1)}}{\partial t^2},$$

and

$$c_{55}v_{1,33}^{(0)} + k_1c_{56}v_{1,3}^{(1)} = \rho\frac{\partial^2 v_1^{(0)}}{\partial t^2} \qquad (12b)$$

$$\gamma_{55}v_{1,33}^{(1)} - 3b^{-2}(k_1c_{56}v_{1,3}^{(0)} + k_1^2c_{66}v_1^{(1)}) = \rho\frac{\partial^2 v_1^{(1)}}{\partial t^2}$$

It may now be noted that the one-dimensional plate equations for generalized plane stress are uncoupled into two sets of equations: (11a,b) governs the extension, width-length flexure and width-shear vibrations and (12a,b) governs the transverse shear and width-twist vibrations. The corresponding deformations governed by these two sets of equations are depicted in Fig. 2a and 2b respectively.

Effect of Electrode Plating

If there are thin metallic electrodes on the faces $x_1 = \pm a$, the mechanical effect of the plating should be incorporated into the plate equations (11) and (12).[7,8] Let upper and lower platings have thicknesses 2a' and 2a'' and densities ρ' and ρ'' respectively as shown in Fog. 3. The stress equations of motion for the platings have the same form as those for the plate (5) and (6). For upper plating

$$\sigma_{13,3}^{(1)'} + H_1^{(0)'} = 2b\,\rho' \frac{\partial^2 v_1^{(1)'}}{\partial t^2}, \text{ etc.} \tag{14}$$

where

$$H_1^{(0)'} = \frac{1}{2a'}\left[T_{11}^{(0)'}(a') - T_{11}^{(0)'}(-a')\right], \text{ etc.} \tag{15}$$

An identical set of equations holds for lower plating except that primes are replaced by double primes.

The continuity conditions of stress and displacement at the interfaces require that

$$T_{1j}^{(n)'}(-a') = T_{1j}^{(n)}(a), \quad T_{1j}^{(n)''}(a'') = T_{ij}^{(n)}(-a),$$
$$v_j^{(n)'} = v_j^{(n)''} = v_j^{(n)}, \qquad n = 0,1, \quad j = 1,2,3 \tag{16}$$

To avoid repetition, the following derivations are restricted to only one of a set of governing equations for plated crystal plates, since the remaining can be obtained in the same manner. The stress resultant in the x_1 direction acting on the faces of the crystal plate with platings is defined as

$$H_1^{(0)} = \frac{1}{2a}\left[T_{11}^{(0)'}(a') - T_{11}^{(0)''}(-a'')\right] \tag{17}$$

By using the definitions (6), (15) and the stress continuity conditions given in (16), the above can be expressed as:

$$H_1^{(0)} = H_1^{(0)} + \frac{a'}{a} H_1^{(0)'} + \frac{a''}{a} H_1^{(0)} \tag{18}$$

Substitution of the corresponding stress equations of motion for upper and lower platings into the last two terms of the right hand side of (18), and use of the displacement continuity conditions given in (16) results in:

$$H_1^{(0)} = H_1^{(0)} + \frac{1}{a}\left(a'\sigma_{13,3}^{(0)'} + a''\sigma_{13,3}^{(0)''}\right) - 2b\left(\frac{\rho'a' + \rho''a''}{a}\right)\frac{\partial^2 v_1^{(0)}}{\partial t^2}. \tag{19}$$

Further substitution of (19) into the first equation of (5), yields one of the stress equations of motion for plated crystal plates:

$$\left[\sigma_{13}^{(0)} + \frac{a'}{a}\sigma_{13}^{(0)'} + \frac{a''}{a}\sigma_{13}^{(0)''}\right]_{,3} + H_1^{(0)} = 2b(1+R)\rho\frac{\partial^2 v_1^{(0)}}{\partial t^2}. \tag{20}$$

where $R = (\rho'a' + \rho''a'')/\rho a$ is the ratio of the mass of the platings per unit area to the mass of crystal plate per unit area.[8] It can be seen that in case the thicknesses of the platings are extremely small as compared to the thickness of the plate, i.e. $a'/a \ll 1$, and $a''/a \ll 1$, the contribution to the stresses due to the platings can be neglected. In the same manner, similar results can be obtained for the other governing equations. Therefore the effect of platings is to increase all the inertia terms in (5) by a factor of $1 + R$.

Extension-Flexure-Shear Vibrations

In the present section, the free, coupled vibrations of extension, width-length flexure and width-shear of a crystal plate with symmetric platings at $x_1 = \pm a$ will be studied (Fig. 1). The free-edge conditions at $x_2 = \pm b$ have been accounted for in (1) while the free-face conditions can be satisfied by setting $H_1^{(n)} = 0$, for $n = 0,1$. To satisfy free-edge conditions at $x_3 = \pm c$, one requires

$$\sigma_{33}^{(0)}\Big|_{x_{3=\pm c}} = \sigma_{32}^{(0)}\Big|_{x_{3=\pm c}} = \sigma_{33}^{(1)}\Big|_{x_{3=\pm c}} = 0 . \tag{21}$$

The appropriate form of solution for (11b) in case an alternating and uniform electric field is impressed over the electrode platings can be written as:

$$v_2^{(0)} = b\sum_{i=1}^{3} A_i\alpha_{1i}\sin\zeta_i x_3 e^{j\omega t},$$
$$v_3^{(0)} = b\sum_{i=1}^{3} A_i\alpha_i\sin\zeta_i x_3 e^{j\omega t}, \tag{22}$$
$$v_3^{(1)} = \sum_{i=1}^{3} A_i\cos\zeta_i x_3 e^{j\omega t},$$

where ω is the circular frequency and ζ_1, ζ_2 and ζ_3 are wave numbers for modes predominant in flexure, extension and width-shear, respectively. The dimensionless frequency and wave numbers are defined by:

$$\Omega = \frac{\omega}{\left[\frac{3k_3^2 c_{44}}{\rho b^2(1+R)}\right]^{1/2}}, \quad \bar{\zeta}_i = \zeta_i b, \tag{23}$$

and they are related by the dispersion relation:

$$F(\Omega, \bar{\zeta}) = |a_{ij}| = 0, \quad (a_{ij} = a_{ji}) \tag{24}$$

where

$$a_{11} = \bar{\zeta}^2 - 3\Omega^2, \quad a_{12} = \hat{c}_{34}\bar{\zeta}^2, \quad a_{13} = \bar{\zeta},$$
$$a_{22} = \hat{c}_{34}\bar{\zeta}^2 - 3\Omega^2, \quad a_{23} = \hat{c}_{33}\bar{\zeta}^2 - 3\Omega^2, \tag{25}$$
$$a_{33} = \hat{\gamma}_{33} + 1 - \Omega^2,$$

$$\hat{c}_{33} = c_{33}^{*}/k_3^2 c_{44}, \quad \hat{c}_{34} = c_{34}^{*}/k_3 c_{44}, \quad \hat{\gamma}_{33} = \gamma_{33}^{*}/3k_3^2 c_{44}^{*}.$$

The roots of (24) or the dispersion curves are shown in Fig. 4. For (22) to be the solution of (11b), the amplitude ratios have to satisfy

$$\alpha_{1i} = \bar{\zeta}_i\left[3\Omega^2 - (\hat{c}_{33} - \hat{c}_{34}^2)\bar{\zeta}^2\right]/D_i, \quad \alpha_{2i} = 3\bar{\zeta}_i\Omega/D_i,$$
$$D_i = (\bar{\zeta}_i^2 - 3\Omega^2)(\hat{c}_{33}\bar{\zeta}_i^2 - 3\Omega^2) - \hat{c}_{34}^2\bar{\zeta}_i^4, \quad i = 1,2,3 \tag{26}$$

Substituting of (22) into the stress-strain relation (11a) and imposing boundary conditions (21), one obtains a set of three homogeneous algebraic equations which govern the amplitudes A_i. For these equations to have non-trivial solutions, the determinant of the coefficients must vanish, i.e.

$$|b_{ij}| = 0, \text{ for } i,j = 1,2,3 \tag{27}$$

where

$$b_{1i} = \mu_i \cos(\bar{\zeta}_i \tfrac{c}{b}), \quad b_{2i} = \sigma_i \cos(\bar{\zeta}_i \tfrac{c}{b}), \quad b_{3i} = \bar{\zeta}_i \sin(\bar{\zeta}_i \tfrac{c}{b}),$$

$$\mu_i = \hat{c}_{34} + (\alpha_{1i} + \hat{c}_{33}\alpha_{2i})\bar{\zeta}_i, \quad \sigma_i = 1 + (\alpha_{1i} + \hat{c}_{34}\alpha_{2i})\bar{\zeta}_i, \quad i = 1,2,3. \tag{28}$$

Equation (27) represents a functional relation between the resonant frequency, Ω, the width-to-length ratio of the plate, b/c, and the mass ratio, R, and is the frequency equation for the present problem.

It can be seen from Fig. 4 that all wave numbers, $\bar{\zeta}_i$, are real when $\Omega > 1$ and $\bar{\zeta}_1$, $\bar{\zeta}_2$ are real and $\bar{\zeta}_3$ is imaginary for $\Omega < 1$. In order to have a real equation corresponding to (27) for $0 < \Omega < 1$, one may replace in (22), $\sin \zeta_3 x_3$ by $\sinh \zeta_3 x_3$ and $\cos \zeta_3 x_3$ by $\cosh \zeta_3 x_3$. all the results from (23) to (28) remain identical except

$$\alpha_{13} = -\bar{\zeta}_3 [3\Omega^2 + (\hat{c}_{33} - \hat{c}_{34}^2)\bar{\zeta}_3^2]/D_3, \quad \alpha_{23} = -3\bar{\zeta}_3 \Omega/D_3,$$
$$D_3 = (\bar{\zeta}_3^2 + 3\Omega^2)(\hat{c}_{33}\bar{\zeta}_3^2 + 3\Omega^2) - \hat{c}_{34}^2 \bar{\zeta}_3^4, \tag{29}$$

and

$$b_{13} = \mu_3 \cosh(\bar{\zeta}_3 \tfrac{c}{b}), \quad b_{23} = \sigma_3 \cosh(\bar{\zeta}_3 \tfrac{c}{b}), \quad b_{33} = -\bar{\zeta}_3 \sinh(\bar{\zeta}_3 \tfrac{c}{b}). \tag{30}$$

Resonant Frequencies

For a +5° X-cut, quartz plate as shown in Fig. 1, the x_1-axis coincides with one of the diagonal axes of the quartz and the x_2-axis makes an angle of +5° with the trigonal axis. The values of (rotated) elastic stiffnesses with respect to the x_i axes of the plate are obtained from Bechmann's value[9] with the slight modification by R.T. Smith[10] and are given in 10^{10} Newtons per square meter as follows:

$c_{11} = 8.674$ $c_{12} = 1.495$ $c_{13} = 0.381$ $c_{14} = 1.721$

$c_{22} = 10.738$ $c_{23} = 0.835$ $c_{24} = 0.134$ $c_{33} = 9.351$

$c_{34} = -2.076$ $c_{44} = 5.694$ $c_{55} = 3.694$ $c_{56} = 1.609$

$c_{66} = 6.087$

and

$$c_{15} = c_{16} = c_{25} = c_{26} = c_{35} = c_{36} = c_{45} = c_{46} = 0.$$

Therefore c_{ij} exhibits monoclinic symmetry. The resonant frequencies of the roots of (27) are computed for a range of values of c/b while the width of the plate (2b) and the mass of electrode platings (R=0.5%) are kept at constant values. The result is plotted as shown in Fig. 5 in which the letters E, F, and WS indicate respectively the portion of the frequency branches corresponding to predominantly extensional, flexural, and width-shear modes while the numeral "n" preceeding the letter indicates the n th mode of a particular type. The dots represent the experimental data obtained by Miss D. M. Spears[14] for the +5° X-cut, quartz plates. The same theoretical and experimental results are also plotted in $\Omega c/b$ and b/c coordinates in Fig. 6 since this way of presenting the data has often been used by previous investigators.[11,2]

Algebraic Formulas

As may be seen from the results shown in Fig. 5 or Fib. 6, the dispersion relation (24) and frequency equation (27) are very accurate in predicting the intricate behavior of the frequency spectrum, for they have accounted for the mechanical coupling of the predominant modes in this range of frequency. However for use in design, simple algebraic formulas for predicting extensional and flexural frequencies separately are desirable. To achieve this object, the coupled equations of motion (11a,b) are uncoupled into extensional and flexural equations as follows:

$$c_{33}^{*} v_{3,33}^{(0)} = \rho \frac{\partial^2 v_3^{(0)}}{\partial t^2}, \quad \sigma_{33}^{(1)} = 2b\, c_{33}^{*} v_{3,3}^{(0)}, \tag{31}$$

and

$$c_{2,3333}^{(0)} + \frac{3\rho}{Y_{33} b^2} \frac{\partial^2 v_2^{(0)}}{\partial t^2} = 0,$$
$$\sigma_{33}^{(1)} = -\frac{2b^2}{3} Y_{33}^{*} v_{2,33}^{(0)}, \quad \sigma_{33}^{(0)} = -\frac{2}{3} b^3 Y_{33}^{*} v_{2,333}^{(0)}. \tag{32}$$

Equations (31) are obtained from (11a,b) by neglecting the coupling terms associated with c_{34}^{*}, while (32) are obtained by a similar procedure used by Mindlin and Forray.[12] Of course (31) and (32) are the classical equations for longitudinal bars and flexural beams respectively.

The roots of the frequence equation for (31) for bars with free ends at $x_3 = \pm c$ are:[13]

$$\frac{c}{b} = \frac{n\pi}{2\bar{\zeta}} \tag{33}$$

where n = 1,3,5,... for antisymmetric modes and n = 2, 4,6.... for symmetric modes in x_3. Instead of using the dispersion relation $\Omega = (\frac{\hat{c}_{33}}{3})^{1/2} \bar{\zeta}$ obtained from (31), the coupled E - WS dispersion relation from (24) is used. This relation is obtained by dropping all a_{2j} terms in (24), and for $\bar{\zeta}^2 < 1$, $\Omega^2 < 1$, it reduces aproximately to

$$\Omega = (\frac{\hat{c}_{33} - \hat{c}_{34}^2}{e})^{1/2} \bar{\zeta}. \tag{34}$$

Substituting (34) into (33), one obtains

$$\Omega = \frac{n}{2} (\frac{\hat{c}_{33} - \hat{c}_{34}^2}{3})^{1/2} \frac{b}{c}. \tag{35}$$

or in dimensional form, with $f = \omega/2\pi$ and $R = 0$

$$f = \frac{n}{4c} (\frac{c_{33}^{*} - c_{34}^{*2}/c_{44}^{*}}{\rho})^{1/2} \tag{35}'$$

Equation (35) agrees with the usal form of extensional frequency formula[11] except that the coupling effect through c_{34}^{*} is included in the term c_{34}^{*2}/c_{44}^{*}.

The frequency equation for flexural equations (32) with free end conditions $x_3 = \pm c$ is[13]

$$\tan \bar{\zeta}\frac{c}{b} = \pm \tanh \bar{\zeta}\frac{c}{b}\,, \tag{36}$$

The roots of the above transcendental equations are approximated by

$$\bar{\zeta} = \frac{\pi}{2}\,(n + 1/2)\,\frac{b}{c} \tag{37}$$

where $n = 1,3,5,\ldots$ for symmetric modes and $n = 2,4,6,\ldots$ for antisymmetric modes in x_3. Here again the coupled F - WS dispersion relation is used. It is obtained from (24) by dropping all a_{1j} terms, and reduces, for $\Omega < \bar{\zeta}$ and $\Omega^2 < 1$, to:

$$\Omega = \left(\frac{\gamma_{33}}{3}\right)^{1/2} \frac{\bar{\zeta}^2}{(1 + \bar{\zeta}^2)^{1/2}}\,. \tag{38}$$

Substituting (37) into (38) one obtains a frequency formula in terms of b/c. In dimensional form and for $R \neq 0$, it can be written:

$$f = \frac{\pi}{8}\left(\frac{\overset{*}{\gamma}_{33}}{\rho}\right)^{1/2} \frac{(n + 1/2)^2\, b/c^2}{[1 + (\frac{\pi}{2})^2 (n + 1/2)^2 (b/c)^2]^{1/2}} \tag{38}'$$

Comparing the above with R. A. Sykes' experimental formula for flexural frequencies:[14]

$$f = K\,\frac{\pi}{2}\left(\frac{\overset{*}{\gamma}_{33}}{\rho}\right)^{1/2} \left(\frac{n + 0.5}{2}\right)^2 \frac{b}{c^2} \tag{39}$$

where K is a coefficient determined experimentally and its value depends on n and b/c, one sees that

$$K = \frac{1}{[1 + (\frac{\pi}{2})^2 (n + \frac{1}{2})^2 (\frac{b}{c})^2]^{1/2}}\,. \tag{40}$$

The frequencies calculated by (35) and (38) are compared with results from the coupled equation (27) as shown in Fig. 7. It can be seen that these formulas give good predictions for $\Omega < 1$. As has been shown by Onoe and Jumonji,[15] the calculated frequencies based on classical flexural beam theory are too high as compared to the observed values, even when the rotatory inertia[13] or both rotatory and lateral inertia are taken into account.[1] The reason for (35) and (38) giving a more accurate prediction is that the effect due to the width-shear deformation has been taken into account through the dispersion relations (34) and (38).

Acknowledgements

The author wishes to thank W. J. Spencer, R. A. Sykes, J. H. Armstrong, P. R. Blomster, and F. S. Welsh for helpful discussions, Miss D. M. Spears for the use of her experimental data, and to acknowledge support of the work by Bell Telephone Laboratories, Allentown, Pennsylvania.

List of References

1. W. P. Mason, Journ. Acoust. Soc. Amer., 6, 246 (1935).
2. H. Jumonji and M. Onoe, Electronics and Communications in Japan, Vol. 51-A, No. 3, 35 (1968).
3. R. D. Mindlin, Quart. Appl. Math, 19, 51 (1961)
4. D. M. Spears, unpublished memorandum.
5. M. A. Medick and Y. H. Pao, Jour. Acoust. Soc. Amer., 37, 59 (1965).
6. H. F. Tiersten, Linear Piezoelectric Plate Vibrations, (Plenum Press, New York) (1969).
7. R. D. Mindlin, Progress in Applied Mechanics, pp. 73-84, Macmillan Co., (1963).
8. R. D. Mindlin and P. C. Y. Lee, Int. J. Solids Structures, 2, 125, (1965).
9. R. Bechmann, Phys. Rev., 110, 1060 (1958).
10. R. T. Smith, personal communication.
11. R. A. Sykes, Chapter VI, Quartz Crystals for Electrical Circuits, edited by R. A. Heising, (D. Van Nostrand Co., New York) (1946).
12. R. D. Mindlin and M. Forray, Jour. Appl. Phys., 25, 12 (1954).
13. Lord Rayleigh, Theory of Sound, (Dover Publications, New York) (1945).
14. R. A. Sykes, unpublished memorandum.
15. M. Onoe and H. Jumonji, paper at Ultrasonic Symposium D5, Cleveland, Ohio, (1966). (Obtained through personal communication.)

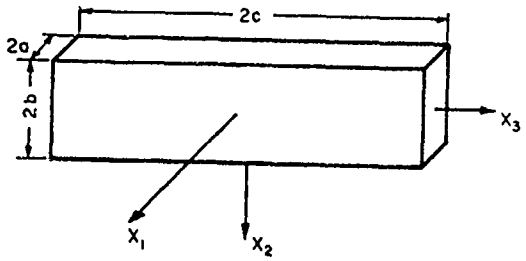
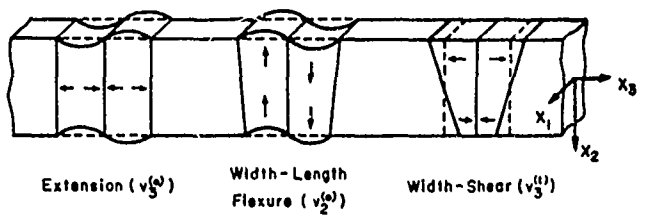
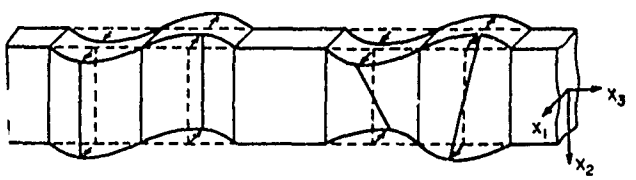


Fig. 1 - Rectangular plate or finite bar with rectangular corss section.



(a)



(b)

Fig. 2 - Plate displacements (a) for extension, flexure, and width-shear motion, (b) for transverse shear and width-twist motion.

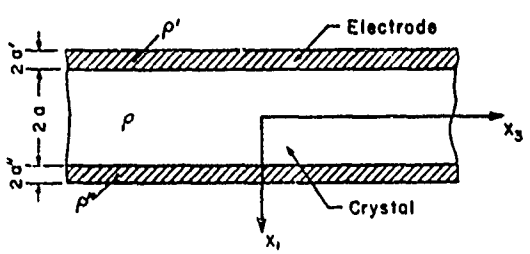


Fig. 3 - Crystal plate with electrode platings.

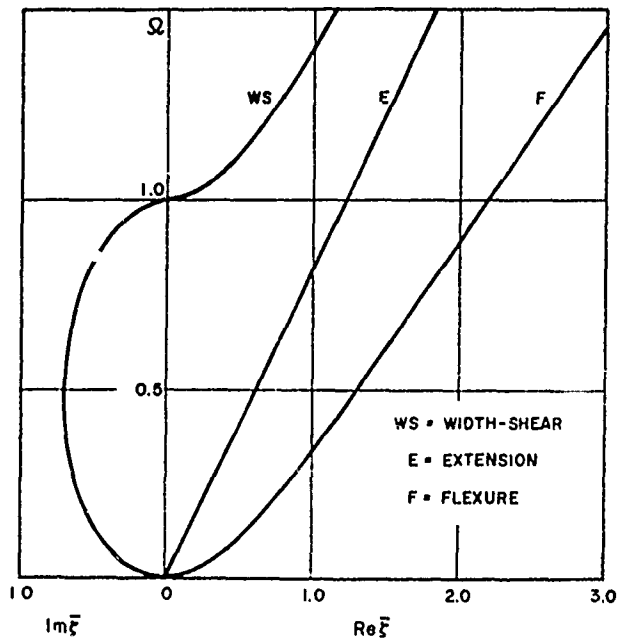


Fig. 4 - Dispersion curves for extensional, flexural, and width shear waves in x_3 direction in an infinite +5° quartz plate. Equation (24).

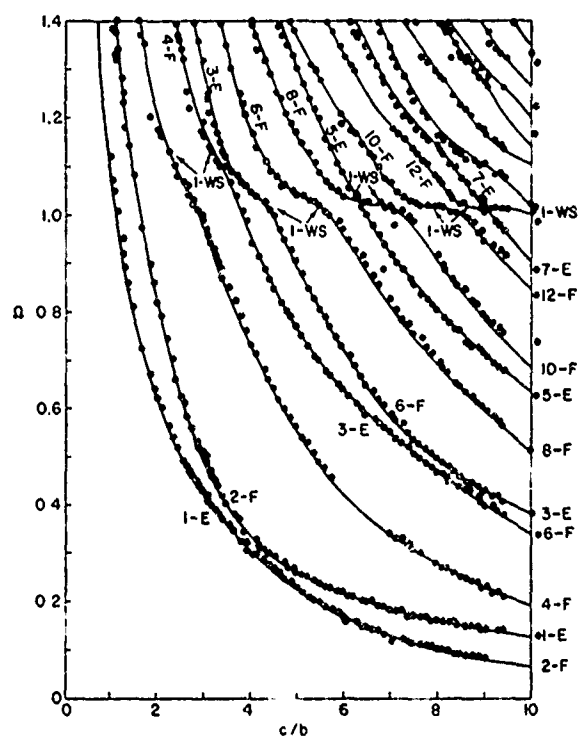


Fig. 5 - Comparison of calculated and measured frequencies of extension, flexure, and width-shear vibration in a rectangular +5° X-cut quartz plate.

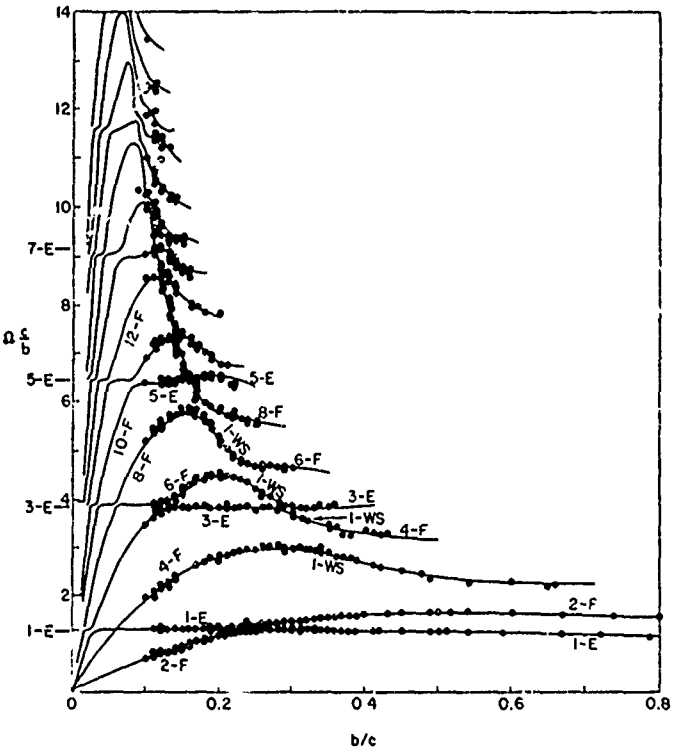


Fig. 6 - Comparison of calculated and measured resonant frequencies of extension, flexure, and width-shear vibration in a rectangular +5° X-cut quartz plate.

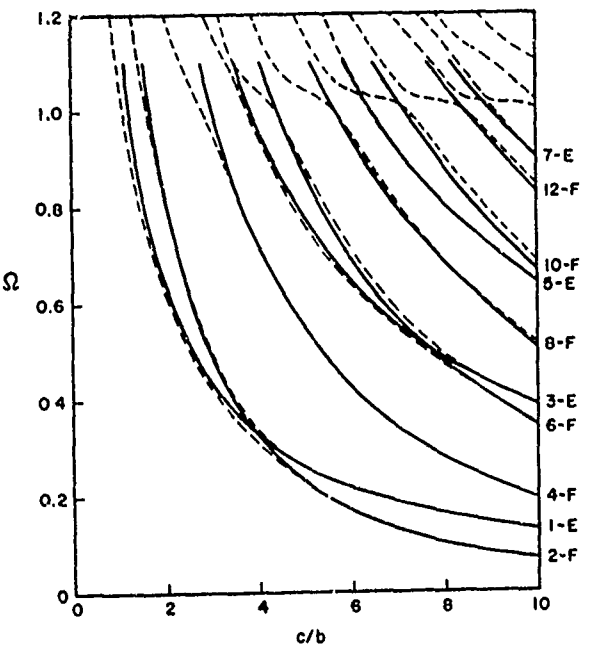


Fig. 7 - Comparison of frequencies calculated by equations (35) and (38) (solid lines) to those computed from coupled equation (27) (dashed lines).

QUARTZ CRYSTAL APPLICATIONS IN DIGITAL TRANSMISSION

Richard B. Hobrock II
Bell Telephone Laboratories, Incorporated
Holmdel, New Jersey

The Bell Telephone System is actively engaged in the development of an integrated, nationwide digital transmission network. The last decade saw the introduction of the first member of the digital network -- the widely acclaimed T1 transmission system. The T1 carrier system provides short haul transmission of twenty-four digitally encoded voice channels on paired cable. Since its inception, over five million voice channel miles of T1 have been installed in the telephone plant. As a result of this success, the Bell System is currently developing a whole hierarchy of digital transmission facilities, ranging in scope from several Kb/s customer data links to several hundred Mb/s coaxial cable, radio and millimeter waveguide systems.

Digital transmission is characterized by three basic operations. First, sampling and coding is the method by which incoming analog signals are converted into a form appropriate for transmission on digital facilities. Second, time division multiplexing is a technique used to combine digital signals from a variety of sources into a single higher speed pulse stream. Finally, regeneration is the process used to reconstruct a digital signal at regular intervals along a digital transmission line. After a brief discussion of each of these operations we will consider the specific role of quartz crystals in the evolving digital network.

Sampling and coding are the two processes which are used to convert analog signals such as speech or television into digital form. The steps associated with this conversion are illustrated in Figure 1. We begin by bandlimiting the analog signal by passing it through a low pass filter. Then we sample the continuous signal at discrete time intervals. If these samples are taken at regular intervals and at twice the highest frequency component found in the signal, then all of the information in the message is contained in these samples. In particular, if the samples are passed through an interpolation filter, the original message will be recovered. This basic principle is known as the sampling theorem and the samples are generally called PAM pulses, standing for Pulse Amplitude Modulation. We have no desire, however, to transmit the PAM pulses because we would have to maintain their exact amplitudes. Instead we perform analog to digital conversion by encoding the amplitude of each sample into a binary word of ones and zeros. This is known as PCM or Pulse Code Modulation. Of course, the length of this word determines how faithfully we can represent the amplitude of each sample. The process of representing a continuum of values with a discrete number of levels is known as quantizing. The error introduced in the process is known as quantizing distortion and is a fundamental signal impairment in digital transmission.

In addition to digital terminals which convert analog signals into digital form, a complete digital network requires time division multiplexing to interface between transmission facilities which transmit digital signals at different rates. The basic structure of a multiplexer and its companion demultiplexer is illustrated in Figure 2. The multiplexer simply combines several pulse streams into a single higher speed bit stream while the demultiplexer performs the opposite task by separating the single high speed stream into its original component parts.

In order to interleave these bit streams, the incoming signals must be synchronous. In the evolving digital network, however, the high speed digital signals will not be synchronized and thus a technique is required to make them appear synchronous. This technique is known as pulse stuffing synchronization. The basic concept requires that the output rate of the multiplexer be slightly greater than the sum of all the incoming rates. Additional pulses are then added to each of the input streams so that the resulting average pulse rate becomes identical to some fixed frequency generated by a crystal clock within the multiplexer. The addition and removal of these extra stuffing pulses is accomplished with elastic stores. These are simply buffer memories, typically storing several bits of information. The READ and WRITE clocks of these stores are independent. In the multiplexer the information bits are written into the synchronizer elastic store under the control of a clock extracted from the incoming signal. This information is then read out from the store at a slightly higher rate, under the control of the local crystal clock. A comparator circuit monitors the number of bits stored in memory. Whenever the storage nears depletion, the comparator demands that a stuff or gap be placed in the READ clock. Of course, the location of this stuff or gap must be signaled to the demultiplexer. This is normally accomplished over a common data channel which is multiplexed along with the data into the output bit stream. The receipt of the synchronization information in the demultiplexer permits the stuffing gaps to be removed from each channel. The destuffed signals are then identical to the original multiplex inputs except for the presence of timing jitter. To remove this jitter the signals are delivered to a desynchronizer elastic store whose READ clock has a frequency which is the average of the jittered WRITE clock. The smooth READ clock is provided by a voltage controlled oscillator, often a crystal oscillator, contained in a phase-locked loop. The loop is required in order to slave the smooth output clock to the jittered input clock.

In addition to synchronization information which indicates the location of the stuffed time slots for each channel, it is also necessary to transmit a framing signal so that the information may be properly demultiplexed. For this purpose a unique code pattern is multiplexed along with the data. Once the position of this unique pattern within the pulse stream has been identified by the demultiplex circuitry, the individual channels may be properly sorted.

The real payoff in digital transmission is the ability to reconstruct a digital waveform which has suffered frequency dispersion and the impairment of additive noise. This reconstruction process is known as regeneration and it is performed at regular intervals along a digital transmission route. The basic functions

associated with a regenerative repeater are shown in Figure 3. These functions are amplification and equalization, timing extraction, and sampling and regeneration. The amplifier-equalizer circuit reshapes the incoming pulse waveform and restores its amplitude so that a pulse-no pulse decision can be made. Since transmission through a length of cable results in severe attenuation of the high frequency content of a pulse waveform, the equalization circuit must in effect boost the relative amplitude of the high frequency components in order to compensate for the distortion introduced by the transmission medium. In practice, a single pulse may be spread over as many as thirty time slots and thus the equalizer performs an extremely important function in reshaping a pulse so that it is spread over few enough time slots to permit regeneration. Some spillover does exist, however, and it is referred to as intersymbol interference.

Following equalization, a timing circuit in the repeater extracts a clock signal from the data stream so that the equalized pulse train can be sampled and subsequently regenerated. This operation is normally performed with a high Q circuit which extracts a sinusoidal component at the timing frequency. This component is then employed to generate extremely narrow sampling pulses which are used to interrogate the data. In some repeaters a monolithic crystal filter is used for timing extraction; in others, a phase-locked loop is often employed with a voltage-controlled crystal oscillator.

The function of the regenerator itself is to reconstruct the received pulse train. This is accomplished by sampling the amplifier-equalizer output with the narrow pulses provided by the timing circuit. When the amplitude of a sample exceeds a specified threshold or decision level, the regenerator produces an output pulse of prescribed amplitude and width; when the sample amplitude is below the threshold, no output is provided.

In practice, a regenerator output is never an exact replica of the transmitted pulse train. The presence of interference on the signal at the instant of regeneration can result in an incorrect decision and thus a digital error. Moreover, phase jitter is introduced on the output due to imperfect timing extraction. Thus in a long repeatered span both digital errors and timing jitter will accumulate. These effects, however, are quite small.

The basic functions illustrated in Figure 3 are common to all regenerative repeaters. In addition, automatic gain control and adaptive equalization are often employed and the transmitted signal usually contains three or more levels. All of this implies a more complicated repeater.

Now that we understand the basic operations associated with digital transmission, it is natural to ask just what are its advantages over conventional analog transmission. First, the availability of silicon integrated circuits permits us to build very inexpensive terminals, both coders and multiplexers. This contrasts with the very expensive and bulky filters associated with FDM or Frequency Division Multiplexing. Second, we require only a very simple repeater plan, whereas analog systems require a family of repeaters differing in their equalization capabilities. Third, through the use of various coders and a hierarchy of digital multiplexers, we can intermix all kinds of signal types for transmission on a single digital facility. Fourth, there exists an obvious advantage in handling signals which are already in digital form, such as data or encrypted speech. Fifth, there is direct compatibility with a switched digital network. And in this decade we will see the advent of time division electronic switching systems. Finally, and perhaps most importantly, once a signal has been placed in digital form, it is virtually insensitive to any further degradation during its tour through the digital network.

A portion of the digital hierarchy planned by the Bell System is illustrated in Figure 4. It contains terminals for coding analog sources into digital signals, multiplexers for getting from one level of the hierarchy to the next, and a variety of digital transmission facilities for carrying these signals. Present analog sources include normal voice channels, the new Bell System PICTUREPHONE®, a 600 voice channel FDM mastergroup, and finally color television. The source of the well known T1 signal is a D1 or D2 channel bank. The D1 channel bank multiplexes 24 voice channels and converts them into digital form. Installation began in 1962 and there are presently over sixty thousand banks in the field. Each unit contains a 1.544 MHz crystal and that adds up to a lot of crystals. The D2 channel bank entered the field last year with more than two thousand banks installed. It multiplexes and converts 96 voice channels into four T1 type signals. Each of these units incorporates a 6.176 MHz crystal as a system clock. At the next level of the hierarchy we find a visual telephone terminal for converting the PICTUREPHONE signal into a T2 digital pulse stream. This coder employs a fundamental crystal operating at the T2 line frequency of 6.312 MHz. And, of course, as PICTUREPHONE service grows, a fantastic number of crystals will be required for this application. At the T3 level of the hierarchy we find two additional coders, one for the 600 voice channel FDM mastergroup and the other for commercial color television. Both of these coders will employ a third overtone 46.3 MHz crystal in their system clocks.

Moving on to consider the multiplexers, we find the M12, which combines four T1 signals into a T2 signal. Each M12 multiplexer uses a 6.312 MHz crystal in its system clock. At the next level we find the M23, which combines seven T2 signals into a T3 pulse stream. The M23 multiplexer contains a 46.3 MHz third overtone crystal for its system clock. At the receiving end, the corresponding M23 demultiplexer will employ seven crystals in the voltage controlled oscillators found in the channel desynchronizers. These will be 6.312 MHz fundamental units. Above the T3 level, we presently have in development M34 and M45 multiplexers. The M34 will combine six T3 signals into a 282 Mb/s T4 signal, while the M45 will combine two T4 signals into a four level 564 Mb/s signal. Fifth overtone crystals in combination with frequency doublers will most likely make up the system clocks for both the M34 and M45 transmitting terminals. In the M34 demultiplexers, 46.3 MHz fundamental crystals may be employed in the voltage controlled oscillators associated with each of the channel desynchronizers.

Quartz crystals find application not only in the digital terminals but in the transmission facilities as well. The basic digital facility is the T1 carrier system which operates on paired cable at 1.544 Mb/s. Digital repeaters are located at one mile intervals with typical lengths ranging from 10 to 25 miles. In 1972 the T2 transmission system will become available, transmitting 6.312 Mb/s over paired cable. Using a new low capacitance cable, repeaters will be located about every three miles with typical lengths ranging from 30 to 200 miles. Each of these repeaters will use an AT cut monolithic crystal filter as a means of extracting a timing wave from the transmitted signal. Two other facilities will carry T2 type signals, in particular, digits on L4 mastergroup and digits on microwave radio.

These existing analog facilities will be used to allow rapid growth of a connected nationwide digital network to permit immediate data transmission and PICTUREPHONE service. The terminals associated with these facilities will of course require crystal clocks.

At the next level of the hierarchy, known as T3, there are no present plans for a transmission facility. Nonetheless there will exist an intrabuilding link at this rate with a 46.304 MHz fundamental AT cut monolithic crystal filter used for timing extraction at the receiving end. Above the T3 level, digital radio and the millimeter waveguide system will most likely carry 282 Mb/s T4 signals. The waveguide facility will simultaneously carry 60 T4 channels, implying a capacity of almost a quarter of a million voice channels. Also in development is a T5 coaxial system carrying a four level 564 Mb/s pulse stream. The repeaters in all of these high speed systems will probably employ phase-locked loops for timing extraction purposes and each of these loops will use quartz crystals in voltage controlled oscillators.

Perhaps the most exciting application of quartz crystals is found in the evolving fully-synchronous digital data system. The realization of a synchronous system calls for the installation of ultra-stable crystal oscillators in cities throughout the country. The digital data system will require synchronization at the T1 level. Data signals can then be synchronously multiplexed up to this level as shown in Figure 5. Customer data rates of 2.4, 4.8, or 9.6 kb/s could be combined by a family of multiplexers into a 56 kb/s signal. Twenty-three of these 56 kb/s pulse stream, along with some housekeeping information, could be synchronously multiplexed up to the T1 level. Thus, in effect, each 56 kb/s data signal would displace one of the digitally encoded voice channels normally associated with the T1 system.

In order to achieve a synchronous data network we first have to synchronize all of the equipment within a single office, and then we have to set about synchronizing all the offices to one another. In order to synchronize the equipment within an office, we establish an office master clock. All system clocks within that office are then slaved to this office master. On a larger scale, all office master clocks are then synchronized to a clock chosen as a system master. These clocks would be phase locked to the master by means of a synchronization network embedded in the system. The phase-locked loops would contain memory, so that in the event of a synchronization outage, the oscillator control signal would be maintained at the same value that existed at the onset of the outage. Thus, in order to minimize the data slips which could be caused by such an unsynchronized clock, we require the use of ultra-stable crystal oscillators.

Let's look at this in some more detail by considering a hypothetical network as shown in Figure 6. We can start by placing a system master in the middle of the country. To this we might slave the regional clocks by means of a linear network. From each regional center we might distribute synchronization information in star-like fashion to the various local offices.

A major drawback in a synchronous network is that provision must be made to allow the network to breathe; that is, the propagation delay of signals through the various transmission media will vary in time, generally as a result of temperature variations. To accomodate the possible differential delay variations between the timing signal and the data stream, we must incorporate buffer memories or elastic stores at the end of every data link. For a slip free nationwide network operating at T1 speeds, these stores become quite large, requiring dozens of memory cells.

Another difficulty associated with a synchronous network is the vulnerability of the system to failures in the transmission links. If we encounter a transmission outage in a timing link, a portion of the network becomes isolated from the system master. For the network shown in Figure 6, the office clock in any node which is no longer receiving timing information serves as a master to all the clocks in the isolated subnetwork. Thus these subnetwork clocks are still synchronized to one another and only traffic between the subnetwork and the main network is endangered. It is our task to minimize this danger.

At the instant an outage occurs, the frequency of a subnetwork master may be offset from the system master due to errors in the synchronization hardware and due to the presence of timing jitter on the synchronization signal. Moreover this frequency deviation can be subsequently aggravated by oscillator drift after the disconnect. The frequency offsets due to both errors and timing jitter can be reduced to quite acceptable levels through proper design of the phase-locked loop. And thus, with the initial frequency offset under control, we need only concern ourselves with the frequency stability of the free running oscillator. Using existing crystal oscillator technology, we believe it is possible to withstand a synchronization outage of several days without causing a slip, that is, without spilling the elastic stores at the end of the data links.

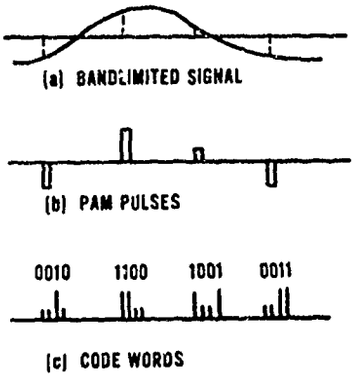


Fig. 1 - Sampling and coding

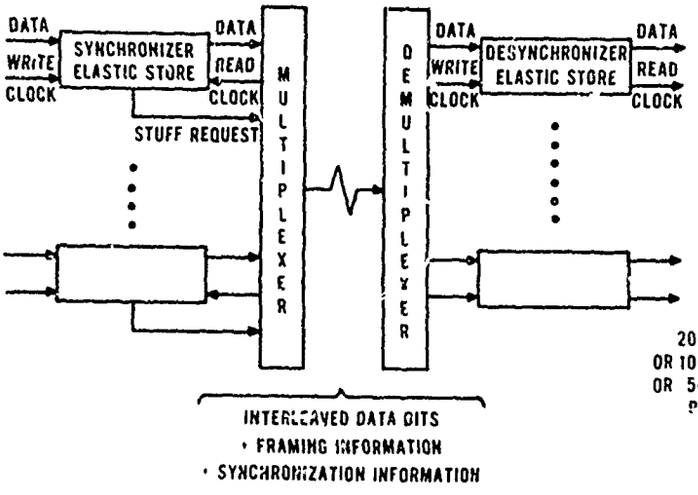


Fig. 2 - Digital multiplexer-demultiplexer block diagram

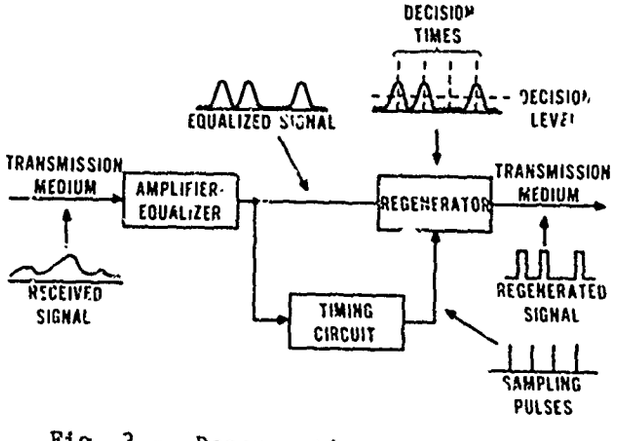


Fig. 3 - Regenerative repeater block diagram

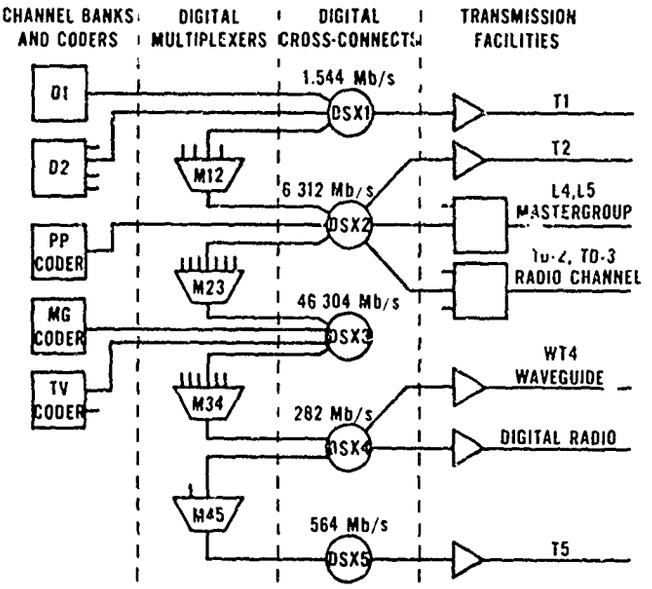


Fig. 4- The evolving digital hierarchy

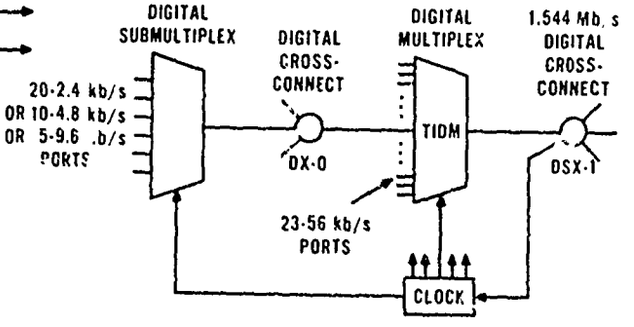


Fig. 5 - The digital data service multiplexing scheme

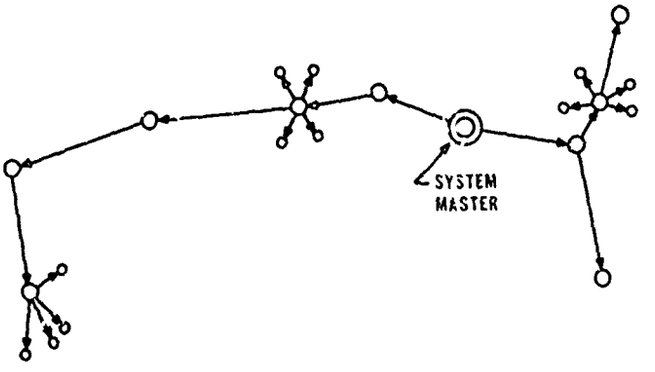


Fig. 6 - A hypothetical synchronous network

FREQUENCY CONTROL DEVICES FOR MOBILE COMMUNICATIOS

By

R. J. Nunamaker
Motorola, Inc.

This paper deals with the impact of the frequency allocations upon the two-way communications industry and the resulting congestion created as a result of this spectrum limiation. It briefly delves into the changes imposed upon the two-way communicar'ons supplier in both equipment and standards to cope with this ever growing demand, and the role that quartz crystal and the many products produced from quartz has played in effectively and economically aiding the designer in making these changes.

To indicate specific applications the F.M. two-way communi ation receiver and transmitter will be analyzed as to the system requirements and the problems created as a result of the congestion and spectrum limitation. Those problems which were capable of solution through the use of quartz products will be further expanded, and the unique manner quartz was used to solve the problems clearly demonstrated.

Other applications to be highlighted are those used in single-sideband systems. It will be further demonstrated that the applications of quartz can be used to correct related single-sideband problems.

THE CRYSTAL-CONTROLLED ELECTRONIC WRISTWATCH SYSTEM: A SI-GATE CMOS-MSI APPROACH

F. H. Musa and R. G. Daniels
Motorola Inc.
Semiconductor Products Division
Central Research Laboratories
5005 E. McDowell Road
Phoenix, Arizona 85008

Summary

This paper describes an electronic wristwatch system which utilizes a crystal-controlled oscillator for a time base. A counter divides the oscillator frequency down by a factor of 2^{16}. The counter, after suitable waveshaping, would drive a miniature motor to turn the watch hands.

A monolithic (MSI) silicon-gate CMOS approach is used for the oscillator, counter, and waveshaping circuitry to achieve low voltage, low power operation. A square wave crystal-controlled oscillator circuit is employed which is relatively insensitive to passive component values and the supply voltage.

The electronics portion of the system (oscillator, 16 stage counter and output waveshaping) is realized on a monolithic 82 × 94 mil chip. Total active device count is 312. Except for an oscillator trimming capacitor, the oscillator passive components are also integrated. Total current drain for the electronics (MSI chip) is less than 4 μA with a 1.3 V battery and a 32 KHz crystal.

Introduction

The block diagram for a crystal-controlled electronic wristwatch system is shown in Figure 1. The oscillator frequency -- controlled by a quartz crystal -- establishes a time base for the watch. This frequency (8 to 131 KHz) is reduced to a few Hertz by the counter. The counter output, after suitable waveshaping and buffering, drives a miniature motor which turns the watch hands. The counter output frequency and necessary waveshaping depend upon the particular motor and gearing arrangement used.

Figure 2 gives typical measured battery data for two (wristwatch size) types of (1.6 V) silver oxide batteries. The same data is given in Figure 3 for two types of (1.3 V) mercury batteries. Notice that the voltage vs. time curves are very flat until just a few days before total discharge -- for all four batteries.

The low voltage, low current restrictions on the electronics for this system are severe; in addition, minimum size and low cost are factors of paramount concern. To satisfy these requirements a silicon-gate CMOS MSI approach was chosen.

Using silicon for the MOS device gate material can reduce the threshold voltages of both n- and p-channel devices to allow operation from a single cell battery and also reduces the device gate to drain and gate to source capacitances for lower dynamic power consumption. The poly silicon gate material may also be used for cross-under connections.

The fabrication details for Si-gate CMOS integrated circuits were described at the 1970 IEEE International Electron Devices Meeting[1] and will not be repeated here. However, the structural details of a Si-gate CMOS IC is depicted in Figure 4 for reference. Figure 5 shows typical complementary device characteristics realized with this process. Threshold voltages in the range 0.6 V to 1.0 V are typical for both P and N channel devices on the same IC chip.

Quartz Crystal

The time accuracy of the watch is dependent on the accuracy and stability of the oscillator frequency. Since a quartz crystal can ensure this type of performance, it was chosen as the frequency controlling element in the oscillator (which is shown schematically in Figure 6). For minimum crystal size the crystal frequency must be made as high as possible. Power dissipation in CMOS integrated circuits is almost linearly proportional to frequency, therefore for extended battery life it is desirable that the frequency be kept as small as possible. It was found that crystal frequencies in the (8 - 131) KHz region satisfy both of these requirements. The NT and XY type crystals cover this range of frequency. Figure 7 shows a photograph of a typical NT type 65.536 KHz crystal whose dimensions are 0.6 × 0.1 × 0.01 inches.

Oscillator Circuit

Figure 6 shows a circuit schematic of the crystal controlled oscillator. It consists of a CMOS inverter, two capacitors, one resistor and a crystal. This circuit configuration was chosen for the following reasons: 1) it is amenable to monolithic CMOS integration (with the exception of the crystal), 2) it operates from very low dc power supply ($V_{DD} \simeq 1.3$ V) at very low current drain, 3) it produces a square wave output whose amplitude is V_{DD} over the frequency range of interest for all reasonable values of C_1 and C_2 (< 50 pf).

Either C_1 or C_2 may be utilized to adjust for exact frequency of oscillation. However, since power dissipation in a CMOS inverter is a function of output capacitance and almost independent of input capacitance, C_1 is utilized for frequency adjustments while C_2 is integrated at a fixed value. The magnitude of the feedback resistance (R_f) may range from 10 MΩ to several hundred MΩ without affecting the oscillator performance (I_{DD}, frequency) appreciably. Due to this insensitivity, R_f may also be integrated.

Theory of Oscillator Operation

The complementary MOS inverter must be biased properly so that enough gain is available to start and to sustain oscillation. This can be done by turning either the n-channel, the p-channel, or both MOSFETS on. It is mandatory, however, that both the n- and the p-channel MOSFETS turn on during one cycle. (This need not take place simultaneously.) DC biasing is achieved as follows: Initially, when V_{DD} is applied, V_1 and V_2 (see Figure 6) are set to ground potential because C_1 and C_2 are much larger than the internal MOSFET capacitances. Q_1, therefore, turns on and Q_2 remains off. Current flows from drain to source in the p-channel device charging capacitor C_2 and finally bringing V_2 to supply voltage potential, at which time current stops flowing. During this transition time, Q_1 might have enough gain to start oscillation, but since g_m drops to zero at zero current, oscillations might not continue due to the relatively short transition time. At this point, V_2 is equal to V_{DD}, and V_1 is at ground potential. Current, therefore, flows through R_f which charges C_1 slowly ($R_fC_1 \simeq .1$-25 m sec) to a positive potential. When V_1 exceeds the threshold voltage of Q_2 this transistor turns on and current flows from its drain to its source which allows C_2 to discharge through Q_2, bringing V_2 toward ground potential. This establishes a positive potential between the source and the drain of Q_1. V_1 increases toward V_{DD} at a much slower rate than V_2 decreases toward ground. If V_{DD} is greater than $V_{T_n} + V_{T_p}$ (here, V_T is defined as the gate voltage at which the channel begins to conduct), Q_1 will eventually turn on and equilibrium will be achieved when both Q_1 and Q_2 have equal currents at which time V_2 and V_1 will be equal. Figure 8 shows how V_1 and V_2 vary from the time V_{DD} is applied. Oscillation can therefore start any time after t_1 provided that the CMOS inverter has enough gain to start and to sustain oscillations. Figure 9 is a photograph showing the biasing level and also the start of oscillation.

Several techniques may be used to determine the necessary and sufficient conditions to initiate continued stable oscillations. By utilizing the MOSFET low-frequency small signal equivalent circuit shown in Figure 10, the oscillator circuit shown in Figure 6 can be drawn as shown in Figure 11. The short circuit admittance (y) parameters for this circuit are:

$$y_{11} = sC_1' + y + g_f$$

$$y_{22} = sC_2' + y + g_o + g_f$$

$$y_{21} = g_m - g_f - y$$

$$y_{12} = -(y + g_f)$$

where $g_f = 1/R_f$, $g_o = 1/R_o$,

$s = (\sigma + j\omega)$ = complex frequency

y = admittance of the crystal in parallel with C_3

$$y = \frac{s^3L_1C_sC_p + s^2RC_sc_p + s(c_p + c_s)}{S^2LC_s + sRC_s + 1}$$

where $c_p = c_o \parallel c_3$,

and the circuit determinant $\Delta y = y_{11}y_{22} - y_{12}y_{21}$ can be expressed as a function of the complex frequency s ($s = \sigma + j\omega$). If Δy is set equal to zero, then a necessary and sufficient condition for oscillation is that $\Delta y = 0$ should have one pair of conjugate roots with positive real components.

$\sigma > 0$ (starting condition for oscillation)

ω = frequency of oscillation (in radians/sec).

Only one such pair of roots should exist in order to guarantee only one frequency of oscillation.

Oscillator Performance

The most important electrical performance parameters of the oscillator are: power dissipation, frequency stability with temperature and supply voltage, the ability to remain operational over the specified temperature range (typically -20°C to +60°C), and the ability to start over a reasonable temperature range (10°C to 50°C).

Unless otherwise specified all data were measured utilizing a typical 32 KHz xy-type crystal whose parameters appear in Table I. Also, threshold voltage V_T as shown in Figures 12, 15, and 16 is defined as the gate voltage which is required to produce a 5 μa channel current when the drain to source voltage is 1.5 volt.

Table I

f_s (KHz)	C_1 (pf)	L_1 (Kh)	C_o (pf)	R_s (KΩ)	Q
32.767	~.0038	~6.25	2.0	30	40,000

Figure 12 shows the current drain of a 32 KHz oscillator operating from 1.3 and 1.6 volts vs. the sum of the threshold voltages of the n- and p-channel devices.

It is essential that the electrical performance of integrated circuits be made as insensitive to variations in the integrated components' magnitudes as possible. Figures 13 and 14 show how current drain and the frequency of oscillation vary with the feedback resistance R_f. It is apparent from these two figures that the performance remains almost the same for $20 < R_f \leq 1000$ MΩ.

Figure 15 shows how the frequency varies with temperature for several combinations of n- and p-channel threshold voltages. $\Delta f/f$ values less than -50 ppm were typical for most crystals (xy cut) tested over the temperature range 0 to 50°C. Figure 16 shows that the frequency of oscillation may be adjusted by approximately 100 ppm by varying C_1 from 10 to 50 pf. Measurements of frequency variations with voltage showed that the frequency changes by less than 2 ppm as V_{DD} varies from 1.3 to 1.2 V with nominal input and output capacitances of 25 pf. This frequency drift with voltage has very little effect on the watch time accuracy because the batteries discharge totally in just a few days after their voltage begins to drop rapidly (see Figures 2 and 3). A frequency drift of 5 ppm will result in time inaccuracy of approximately 13 sec/month.

Frequency Divider

The frequency divider (or counter) for this system consists of a cascade of (binary) toggle flip-flops each of which divides the frequency by two and has a constant power/frequency ratio of ~ 25 nW/KHz when operated from 1.3 - 1.6 V supply. Figure 17 shows the total current drain vs. the input frequency for a Si-gate CMOS IC counter at 1.6 and 1.3 volt supplies. (The maximum input frequency at 1.3 V is about 1 MHz.)

IC Chip

Figure 18 is a die photograph of a Si-gate CMOS electronic wristwatch IC. The size of this chip is 82 x 94 mils; the total active device count is 312. The oscillator inverter, output capacitance (≈ 20 pf - MOS) and feedback resistance (≈ 50 MΩ) are integrated on the chip. The oscillator input capacitance (frequency trimmer) is connected externally. The frequency divider consists of 16 flip-flops. The outputs of the last few flip-flops are connected to two Nand gates to produce two low duty cycle pulses which are out of phase. These pulses drive the two power inverter buffers (2 large P channel devices and 2 large N channel devices). The buffers drive the motor. Because of the low duty cycle, the average motor current is reduced to an acceptable level. The total current drain for this IC is typically less than 4.0 microamperes with a 1.3 V mercury battery, and a crystal with characteristics similar to those given in Table I.

Acknowledgement

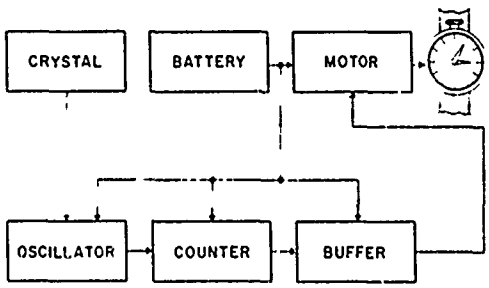
The authors wish to acknowledge R. R. Burgess, E. M. Cheskis, U. S. Davidsohn, J. W. Foltz, T. R. Huffman, J. B. Price, J. L. Rutledge, and A. G. Thiele for their outstanding contributions during the development of the Si-gate CMOS technology and to J. R. Preece for artwork preparation.

References

1. "Silicon-gate CMOS IC's" by R. R. Burgess and R. G. Daniels, presented at the International Electron Devices Meeting in Washington, D.C., October 1970.

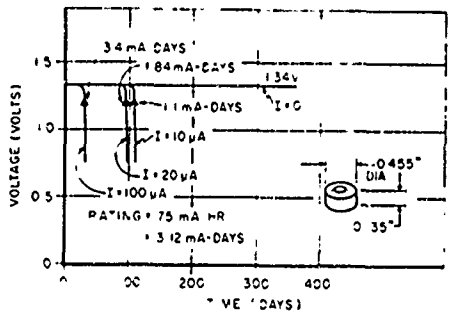
2. "The Electronic Wristwatch: An Application for Si-gate CMOS IC's", by R. G. Daniels and R. R. Burgess, presented at ISSCC in Philadelphia, Pa., February 1971.

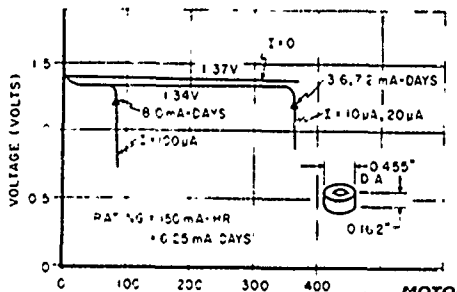
ELECTRONIC WRISTWATCH BLOCK DIAGRAM



MOTOROLA
Semiconductors

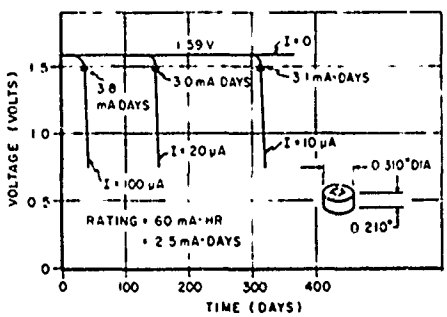
MERCURY BATTERY LIFE

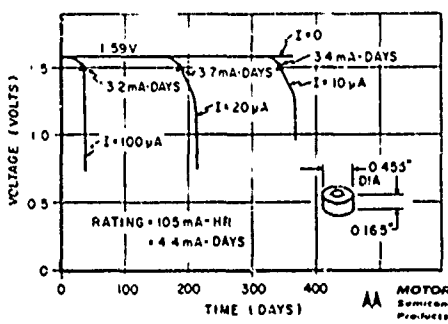




MOTOROLA INC.
Semiconductor
Products Division

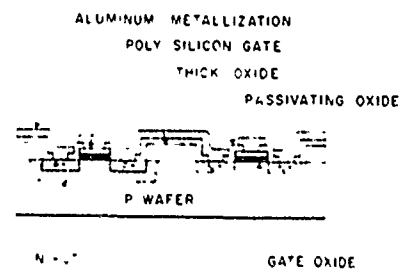
SILVER OXIDE BATTERY LIFE





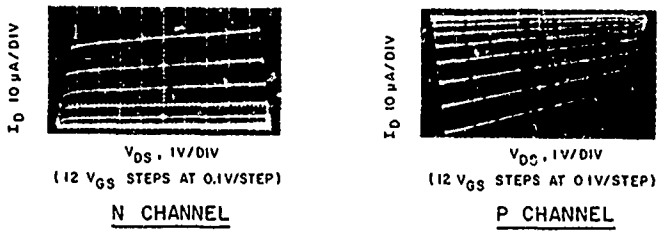
MOTOROLA INC.
Semiconductor
Products Division

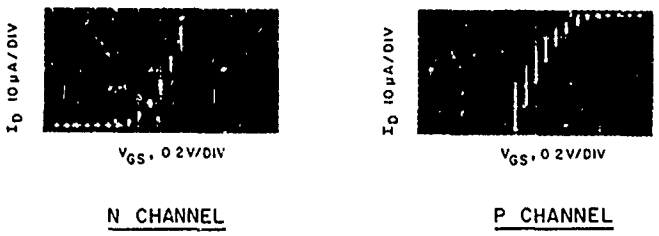
FINAL Si-GATE CMOS IC STRUCTURE



MOTOROLA INC.
Semiconductor
Products Division

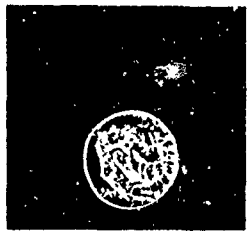
TYPICAL SILICON GATE MOS CHARACTERISTICS

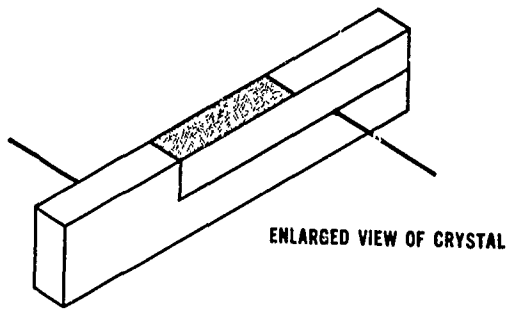




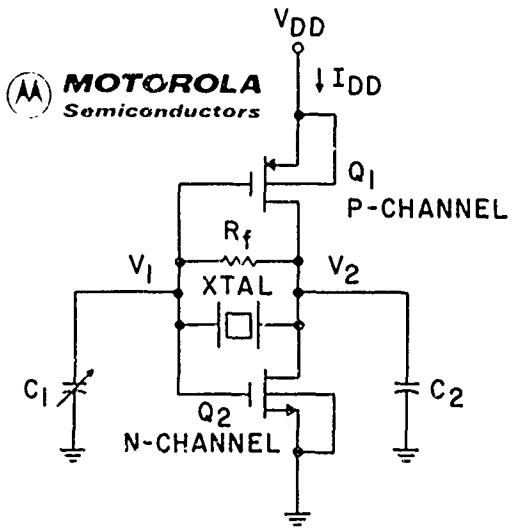
MOTOROLA Semiconductors

MOTOROLA Semiconductors



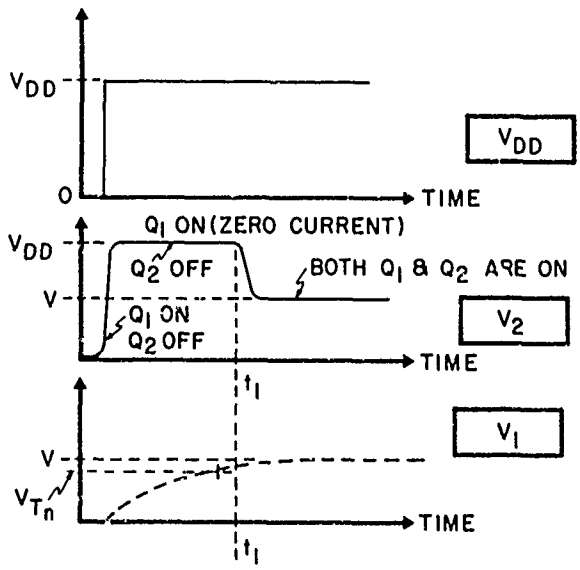


QUARTZ CRYSTAL



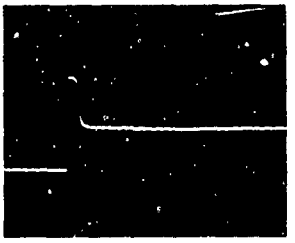
CRYSTAL CONTROLLED OSCILLATOR

MOTOROLA INC. Semiconductor Products Division



DC BIASING OF OSCILLATOR

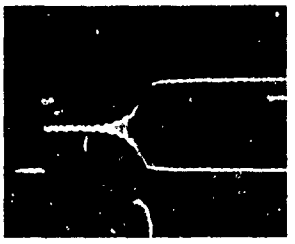
UNIT No. 27 W 50



V_2 OUTPUT VOLTAGE

2 nS/cm

C_1 = 25 pF
C_2 = CMOS (INTEGRATED)
R_f = 40 MΩ

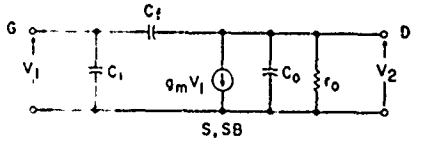


V_2 OUTPUT VOLTAGE

50 mS/cm

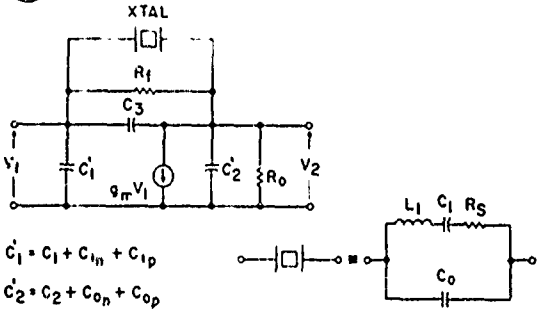
START OF OSCILLATION

MOTOROLA INC. Semiconductor Products Division



MOSFET LOW FREQUENCY SMALL-SIGNAL EQUIVALENT CIRCUIT

MOTOROLA INC. Semiconductor Product Division



$C_1' = C_1 + C_{in} + C_{tp}$

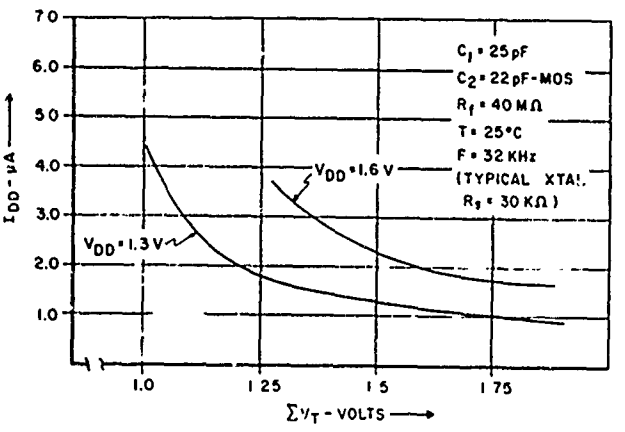
$C_2' = C_2 + C_{on} + C_{op}$

$C_3 = C_{fn} + C_{fp} + C_{PARASITIC}$

$g_m = g_{mn} + g_{mp}$

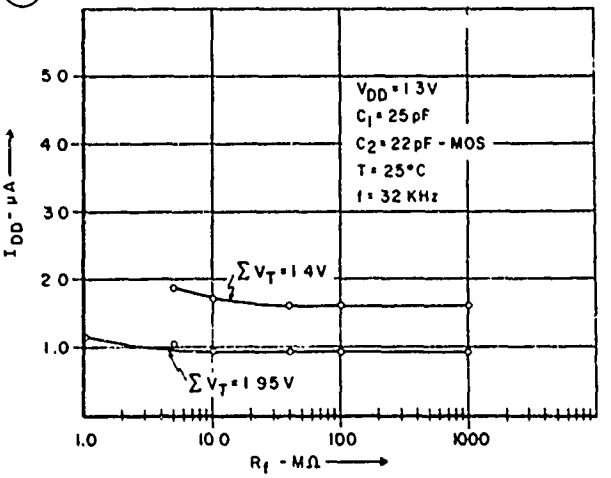
OSCILLATOR SMALL-SIGNAL EQUIVALENT CIRCUIT

MOTOROLA INC. Semiconductor Products Division



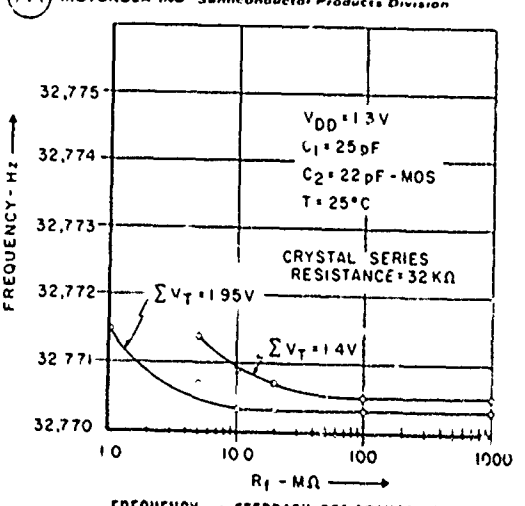
I_{DD} vs SUM OF THRESHOLD VOLTAGES

MOTOROLA INC. Semiconductor Products Division

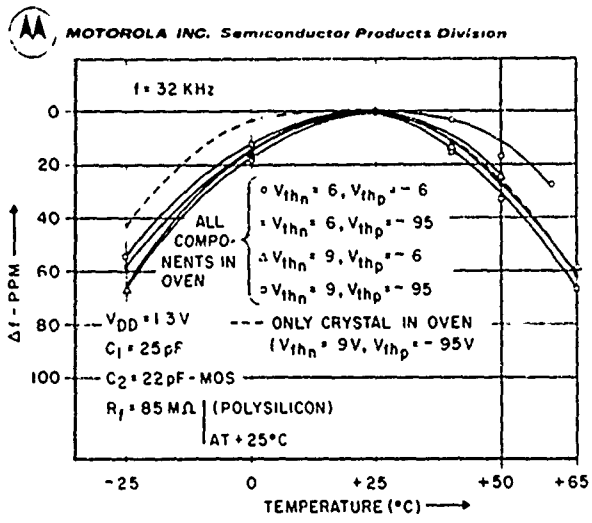


I_{DD} vs FEEDBACK RESISTANCE (R_f)

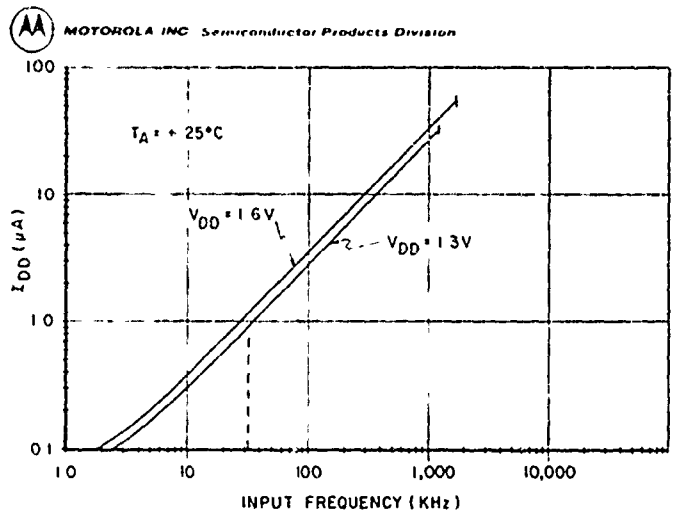
MOTOROLA INC. Semiconductor Products Division



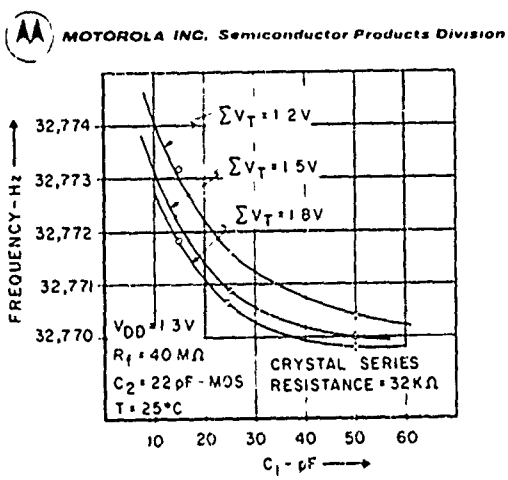
FREQUENCY vs FEEDBACK RESISTANCE R_f



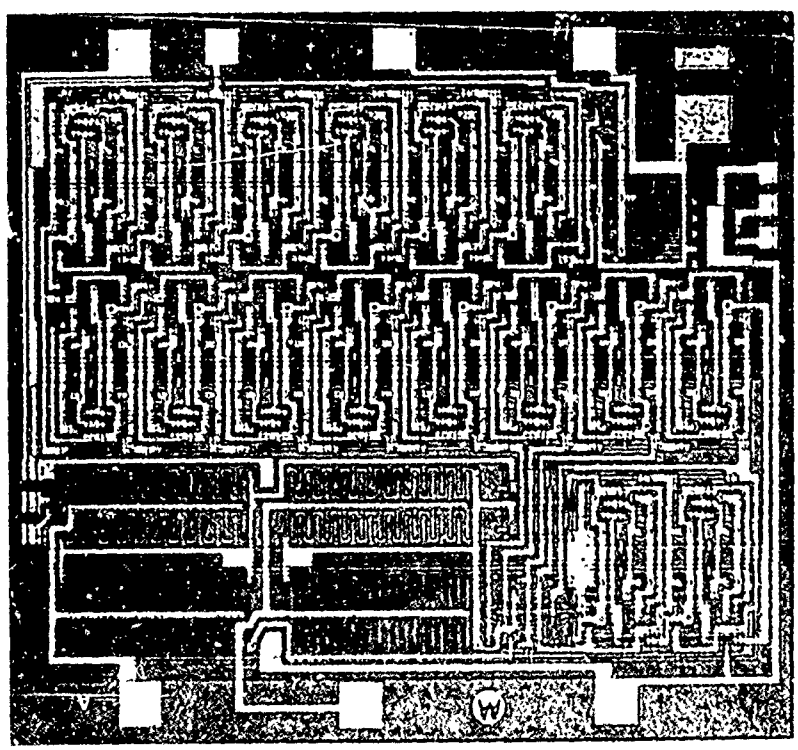
FREQUENCY VARIATIONS vs TEMPERATURE



TYPICAL CURRENT DRAIN vs FREQUENCY FOR Si-GATE CMOS COUNTER



VARIATIONS OF FREQUENCY WITH INPUT CAPACITANCE



FREQUENCY AND TIME IN AIR TRAFFIC CONTROL AND COLLISION AVOIDANCE APPLICATIONS

Vernon I. Weihe
Electronic Systems Consultant

Introduction

Precision frequency control with highly stable clocks plus new information system devices will be combined, during this decade, to solve a number of the more difficult problems of Air Navigation and Traffic Control. Precision control already has been applied to the routine operation of several ground based aids (Loran, Decca, and Omega), making ground station synchronization less dependent on the stability and continuity of radio links between stations. This application will be covered later in this symposium.

A somewhat more recent application is that in the Time/Frequency Collision Avoidance System.[1] This system has been developed and flight tested in a form which specifies a minimum acceptable stability of 1×10^{-8}, and a clocking accuracy within two microseconds of the master. The radio frequency and the system time are based on the zero magnetic field cesium atomic resonance, defined as 9,192,631,770 Hz.[2,3,4]

Air Traffic Control

The Air Traffic Control System is a global complex composed of a variety of communication, navigation, radar and other systems, representing many generations of hardware. In common civil/military unrestricted airspace around the world, the system provides a traffic separation service and smooths traffic flow for all types of aircraft. Operating procedures and separation criteria for the global system are established by the International Civil Aviation Organization. These are upgraded in certain high traffic density States to suit National needs.

The ATC system slowly evolves under the diverse leadership of many groups within Government and Industry. At no time is there a universally accepted long term master plan. The system's great size and complexity leave little opportunity for major changes of a revolutionary nature. Manpower and financial support are normally lacking to achieve speeded obsolescense of existing avionics and ground systems to accommodate rapid introduction of the radically new.

Hence, it should come as no surprise that few of the revolutionary achievements of the past decade have found their way into the routine operations of the system. Some new componentry of the sixties has been assimilated, but the fundamental system organizational plans and configurations in concept predate the information systems era.

Personnel and organizational changes have been made recently which appear to be improving the quality and posture of ATC systems engineering management. The decade of the seventies under this new leadership and having the new electronics available should be a period of exceptional progress.

The ATC system should be updated to include:

1. Better data acquisition.
2. Increased use of Information Systems.
3. Adoption of a data link system to augment voice radio.
4. Implementation of Collision Avoidance Avionics.
5. Expanded use of Area Navigation.
6. Application of satellites and
7. Introduction of a new microwave scanning beam landing guidance system.

In all of these systems precise frequency control has a significant role to play.

ATC Data Acquisition

Since the origin of air traffic control, methods for the acquisition of data on aircraft position, altitude and identity have always been weak. It is in this area that new precise frequency control technology may make its greatest contribution.

Voice position reports (at times over very noisy radio circuits), ground based primary radar and radar beacons are utilized. All of these have major deficiencies, and none of them fits well into a global automated system concept.

In the current system and in new elements, being readied for use, inherently low bandwidth functions are being accomplished with non-essential intermediate human interfaces and with surveillance, rather than control, concepts and technology. For close-in air terminal regions the deficiencies are not unacceptably severe but for transition and, more notably in the enroute zone, the weaknesses are very significant.

In the high altitude enroute zone, surveillance type automation devices are inserted between the aircraft system and air traffic control. These operate independently and primarily process wideband noise interspersed with some useful data. The system should process many useful signals in a small field of noise. Because of the independent nature of the ground stations, the useful signals are processed in random groups fundamentally disorganized and highly redundant. Thus the system is less reliable and more costly than one based on more rational concepts. Then, too, the system is landlocked while most of the earth's airspace is over the oceans.

With state of the art frequency control and clocking, it is feasible to provide aircraft identity,

position and altitude in single systematic queues for any airspace, over land or sea. The data can be organized so that one and only one signal is transmitted in a given time-slot and with a high ratio of signal to noise. Noise created within the system itself can be essentially eliminated. Such a system concept is compatible with buffers of conventional information processing systems, is more reliable and cost effective. It also can conserve radio bandwidth.

Data Acquisition "Strawman" No. I. Let us assume that all aircraft above the 18,000 foot level carry the new T/F Collision Avoidance System. Each aircraft will then transmit identity and altitude in its own time slot once in each three second EPOCH. No signals by other aircraft will be transmitted in its time interval.

Any synchronized ground station receiver within line-of-sight range can then output identity, altitude and geographic position for all aircraft, in a sequence which is systematic, jam free, non-redundant and has a high signal to noise ratio. No ground based interrogations are required.

The T/F CAS system operates in the 1600 MHz region and propagates master fine synch signals from a single heirarchy 0 Master Ground Station in each region. An integrated avionics relay propagates time of lower heirarchy to remote airspace over land and sea. The clock steps down slowly in 0.05 microseconds steps to heirarchy 40, the two microsecond (1966 feet one way transit time) limit.

Radio frequency service allocations above and below CAS are reserved for the mobile (up and down) links of satellite communications relay system. By proper organization CAS and the L-Band satellite system can be linked to provide a global all airspace precision data acquisition and time reference system.

Programming such a system involves:

1. The use of space deployable wide aperture satellite antennas,
2. Some communications traffic engineering, and
3. The allocation of CAS and satellite channels which use the same radio frequency synthesizer circuits.

Data Acquisition "Strawman" No. II. For more than a decade the inherent deficiencies and the limited traffic capacity of the Air Traffic Control Radar Beacon System have been well documented. Because of exhortations not to tamper with the system relatively weak pleas are made for a means whereby interrogators trigger only those aircraft which are relevant. There are more than 60,000 transponders in use; the planned phase out of primary radar signal processing (for improved system economics) will drastically increase that number. Over interrogation provides an excessive number of replies to be processed, and both the signal quality and the integrity of the replies are degraded. Deterioration is greatest in areas of greatest need.

With precise timing techniques now available a transponder can be adapted to transmit identity and altitude standard codes on a time schedule in lieu of interrogation. An individual ground station, without transmitting interrogating signals, could obtain identity, altitude by code, range (via one way ranging) and azimuth (via instantaneous D.F.). Nearby ground stations netted to use range only could obtain position via tri-lateration (Rho/Rho/Rho) using high gain omnidirectional antennas on 1090MHz.

The processing environment again would deal with one and only one transmission from each aircraft in each time slot and the transmission sequence would not be repeated until the next time EPOCH.

The real world side of the processor buffer would queue aircraft on a time slot by time slot basis. The first signal arriving in each time slot could be used and the remainder "dumped" to avoid the need for a larger processor or to prevent "packing" an existing one.

In a CAS equipped aircraft the timing control could be supplied within the CAS/ATC/RBS system. With data link an inhibit signal could reduce slots used in an area to that required by ATC.

Casual analysis will show that a number of compatible system design options exist within the above concepts.

Satellite Applications[5]

In the near term, satellite communications (probably in L-Band) will augment, then later supplant, HF Radio for air/ground communications on long flights outside continental U. S. A. Automatic position reporting to ATC and to operations will become available. ATC and operational messages will be communicated. Most, if not all of the traffic will be in digital data format, serving information systems in the aircraft and on the ground.

Because of recent large investments, by air fleet operators, in inertial and doppler radar navigation systems and the ability to update them with Vortac, Loran and Omega, there is little current enthusiasm for satellite navigation. However, near term developments of the seventies may modify this attitude.

Phase locked digital data communications systems with precise control of frequency and time inherently have within them the ability to provide accurate geographic lines of position. They also can provide excellent means for the global propagation of very precise time to aircraft, ships and fixed or moving points ashore.

Global Timing and Omega

Omega propagates signals which suffer little attenuation as they transit Earth, held between its surface and the ionosphere. The Omega commutation scheme (Figs 1, 2) is unambiguous and synchronized to UT-2 with the 10 second period beginning at 0000 hours and repeating at 10 second intervals. Thus Omega is not only a navigation system for air, surface and submerged craft but can serve as a worldwide timing system.

The basic Omega radio frequency (10.2 KHz) has a wavelength of 16 nautical miles. With C.W phase comparison an ambiguity occurs each eight miles along the baseline between stations. A second frequency (13.6 Khz) is beat with the basic to obtain 3.4 KHz which is ambiguous each 24 miles. Still a third (±1 1/3 KHz) is beat to move the ambiguous lanes out to 72 nautical miles.

Position uncertainties can be removed by continuous counting of lanes from point of departure, by DR, or by obtaining a coarse fix from other aids to navigation. But accurate time from complementary systems also can be utilized.

In the recent CAS flight test program the master ground station was time locked to the nearest Loran-C master. A global master timing system to serve the CAS system is not yet programmed.

When the CAS master stations are tied together via satellite or other means all airborne stations can be within two microseconds of CAS master time. By knowing the exact times of transmission of Omega signals and having precise time aboard, continuous Omega position fixes can be obtained on the basis of the one-way propagation time from each station to the user. Radio waves propagate at 6.18 microseconds per mile. For lane identification:

N. Miles	Microseconds
8	49.4
24	148.3
72	445.0

Thus using either CAS or Satellite timing ships and aircraft have a new means to resolve lane ambiguities, or alternatively a means for resolving lane ambiguities for a single frequency simplified Omega receiver. For example, a simple Search and Rescue Omega receiver of the single frequency type (10.2 KHz) could obtain lane resolution via a timing source, automatically, prior to the emergency incident and use Omega alone in the emergency mode.

A shipborne Omega with a good companion clock could receive time infrequently from any CAS equipped aircraft within line of sight range or from any satellite which is time referenced.

Thus when CAS time and synchronous satellite time are used within a worldwide standard data acquisition system for ATC, navigation, communication, collision avoidance and search and rescue are also served.

In figures 3, 4, and 5 coverage for synchronous satellites for several angles above the horizon are shown. Probable Omega Station locations and some of their baselines are shown in figure 5. Coverage for the high latitudes can be supplied by satellites in random Polar or Sun Synchronous orbit.

More Conventional Requirements for Precise Radio Frequencies and Stable Clocks

VHF Radio and Data Link

Much of the progress made in Air Traffic Control Systems to date has come from advances in frequency control. The "King George" VHF voice radio communications airborne set of World War II was channelled at 180 KHz. In the postwar period only those channels compatible with 200 KHz were retained for service. Successive equipment generations used 100 KHz, then 50 KHz. The current trend is toward 25 KHz.

Since the system is still double side band AM, and still carries only a 3 KHz voice service, it is apparent that the frequency control community has not yet exhausted all opportunities to improve the system.

In the seventies the ATC Air/Ground Communications System should be connected to National and International ground based data networks with automatic switching of voice circuits (e.g. AutoVon) and automatic digital message switching (ADMS of AutoDin). Despite more than a quarter century of discussion, planning and aborted development, the air/ground voice system even now lacks a dial (or touchtone) and does not have a telephone bell. Although many aircraft have several channels to the cockpit, crews are forced to listen continuously to intercept calls.

Avionics Radio Frequency Planning for the Seventies

ICNI and Communications Traffic Engineering

Aviation is now at least five years into the planning of new integrated avionics systems. Yet radio frequency allocation planning is far behind the hardware engineering. Integrated communication, navigation and information (ICNI) systems cut across the entire gamut of ATC systems.

In ICNI central digital modular processing time-shares wideband multiplexed electronics, single coaxial cables replace large bundles of copper wire, discrete analog and digital units of the past are no longer identifiable except as sensor, control and display points. A decrease in dependence on line-of-sight limited propagation paths also is planned.

But the line-of-sight limitation can be a 'blessing'. It allows the same radio frequency channel to be utilized again and again. As yet the sophisticated type of communications traffic engineering required to implement the non line-of-sight goal is singularly lacking.

In this activity junior systems planners, not having learned of the "momentum of money" (i.e., the long tenure of large investments in worldwide standardized hardware) seek rapid adoption of incompletely defined revolutionary systems, while their more senior partners are too insensitive to the new revolutionary electronics. They emphasize patch on top of small patch, and devote their efforts to shoring up obsolescent systems of a bygone era. Bygone by only a decade, but what a decade. Between the two generations of engineers lie many viable courses for truly constructive action.

Radio spectrum is the priceless "real estate" upon which most avionics systems are built. Wise use of it is essential. Poorly planned spectrum produces poorly configured airborne systems.

A current challenge, not being met comprehensively, is that of frequency channelling the bands and services listed below, so that a single modular radio frequency standard and set of synthesizers can perform all of the functions in a way which is ideally compatible with ICNI philosophy. For aviation this challenge, properly met, could open the way for more progress than any since the ITU Conference at Atlantic City, (1947).

In the postwar period without sound operational and system judgment, aeronautics opted for VHF, AM voice radio and the maritime chose FM voice radio in the same bands. For two decades they haven't been able to communicate. Operations of both have suffered. In these new bands it's hoped that similar blunders will be avoided.

New developments requiring careful channeling are in the following bands:

1435-1537.5 MH3 Maritime Mobile
1537.5-1542.5 MH3 Aeronautical Mobile
Marine Mobile
1542.5-1557.5 MH2 Aeronautical Mobile
1557.5-1637.5 MH3 Aeronautical Radio Navigation
1637.5-1640 MH3 Maritime Mobile
1640-1645 MH2 Maritime Mobile
1645-1660 MH2 Aeronautical Mobile
4200-4400 MH3 Aeronautical Radio Navigation
5000-5250 MH3 Aeronautical Radio Navigation
15.4-15.7 GH3 Aeronautical Radio Navigation

Collision Avoidance, Satellites, Microwave Landing Guidance.

To be more specific from an Avionics and ATC viewpoint, the Collision Avoidance System is really a first urgent use of T/F technique, it's the beginning and not the end of a road. In time the system will be expanded to accomplish additional needed services.

In like manner, as indicated previously, satellite communications, per se, is not an end but a beginning of the use of communications relay. Communications is first because there is an urgent need to place less reliance on high frequency radio.

Next comes the requirement to supplant the World War II fixed beam localizer and glide slope and to replace it with a Microwave Scanning Beam Landing Guidance System capable of category III flare out to touchdown.

From a frequency control and stable clock veiwpoint all of these need to be compatibly organized to minimize redundancies of the 'afterthought' variety and to economically justify high performance levels than that which a group independent systems could utilize. In this area light weight general aviation aircraft must be adequately served.

Some Current and Proposed Allocations for the Services noted above.

1. 1540-1560 MHz, Aircraft reception from satellite
*1460-1570 MHz, ILS Glide Slope
**1595-1625 MHz, Collision Avoidance System
*1630-1640 MHz, Radio Altimeters (secondary)
1640-1660 MHz, Aircraft transmission to satellite

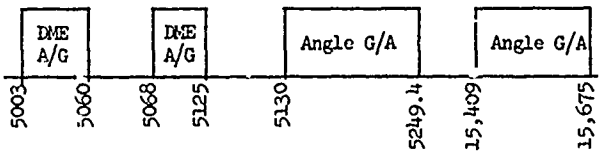
Notes: *Not apt to be in general use.
**F1 1600 MHz
F2 1605 MHz
F3 1610 MHz
F4 1615 MHz

2. Landing Glide Slope Frequency Allocation Plan (See following chart for details)

5003-5060 DME A/G
5068-5125 DME G/A
5130.0-5249.4 Angle C-Band
15,409.0-15,538.1 Angle Ku-Band

3. Search and Rescue (SAR) harmonics of 121.5 MHz are also applicable in 1540-1660 MHz band.

4. LGS Frequency Allocation Plan



Channel	DME A/G	DME G/A	Angle C-Band	Angle Ku-Band
1	5003	5068	5130.0	15,409.0
2	5003	5068	5130.6	15,409.9
5	5003	5068	5132.4	15,412.6
6	5003	5068	5190.0	15,413.5
7	5003	5068	5190.6	15,414.4
10	5003	5068	5192.4	15,417.1
11	5006	5071	5133.0	15,418.0
15	5006	5071	5135.4	15,421.6
16	5006	5071	5193.0	15,422.5
181	5057	5122	5184.0	15,571.0
191	5060	5125	5187.0	15,580.0
195	5060	5125	5189.4	15,583.6
196	5060	5125	5247.0	15,584.5
200	5060	5125	5249.4	15,588.1

Note: This LGS Frequency Allocation Plan was excerpted from RTCA Paper 189-70/SC 117, SFDT-27 dated September 5, 1970, the Signal Format Development Team Report.

Recommendation

In pursuit of the above stated and implied goals, it is recommended that this group, whose members know the most about frequency controls' present status and future potential, help cognizant organizations to make the best possible use of the current state of the art, and meanwhile stake out suitable options for future progress. Some of the cognizant organizations are: 1. FCC and IRAC; 2. FAA; 3. RTCA; 4. RTCM; 5. ICAO; and 6. ITU.

References:

1. Flight Test and Evaluation of Airborne Collision Avoidance System
   a. Summary Report
   b. Vol. I and Vol. II, Final Report
2. Air Transport Time-Frequency Collision Avoidance System, ARINC Characteristic No. 587, Sept. 1, 1970.
3. Airborne Collision Avoidance System Development IEEE Transactions on Aerospace and Electronic Systems Vol. AES-4 No. 2, March 1968.
4. Proceedings of 24th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Fort Monmouth, New Jersey.
5. Journal of the Institute of Navigation, Vol. 17, No. 3, Fall 1970.

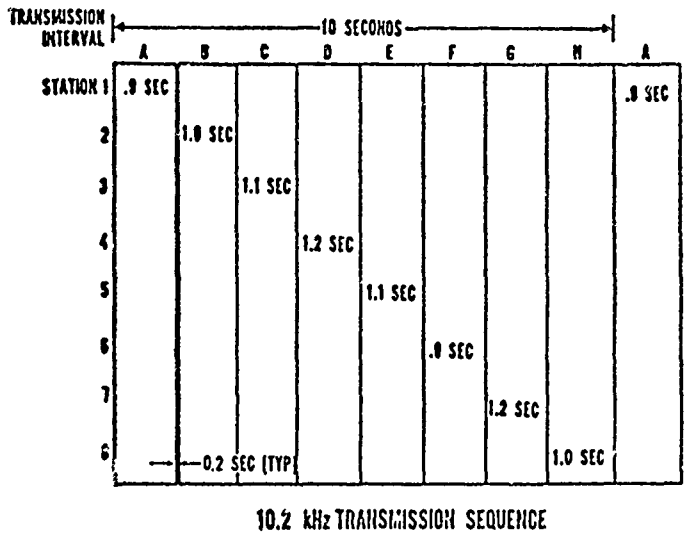


FIGURE 1 SIMPLIFIED 10.2 KHz TRANSMISSION FORMAT.

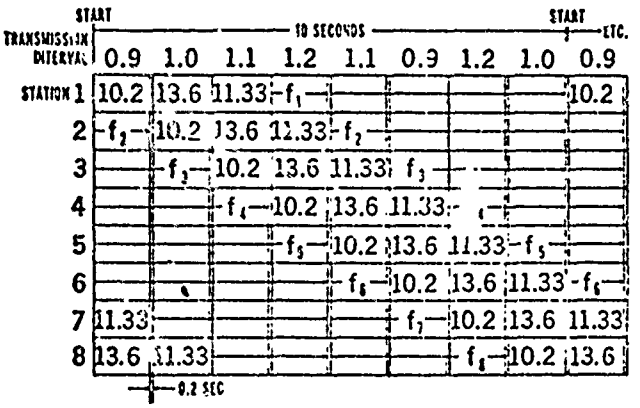


FIGURE 2 OMEGA SIGNAL FORMAT.

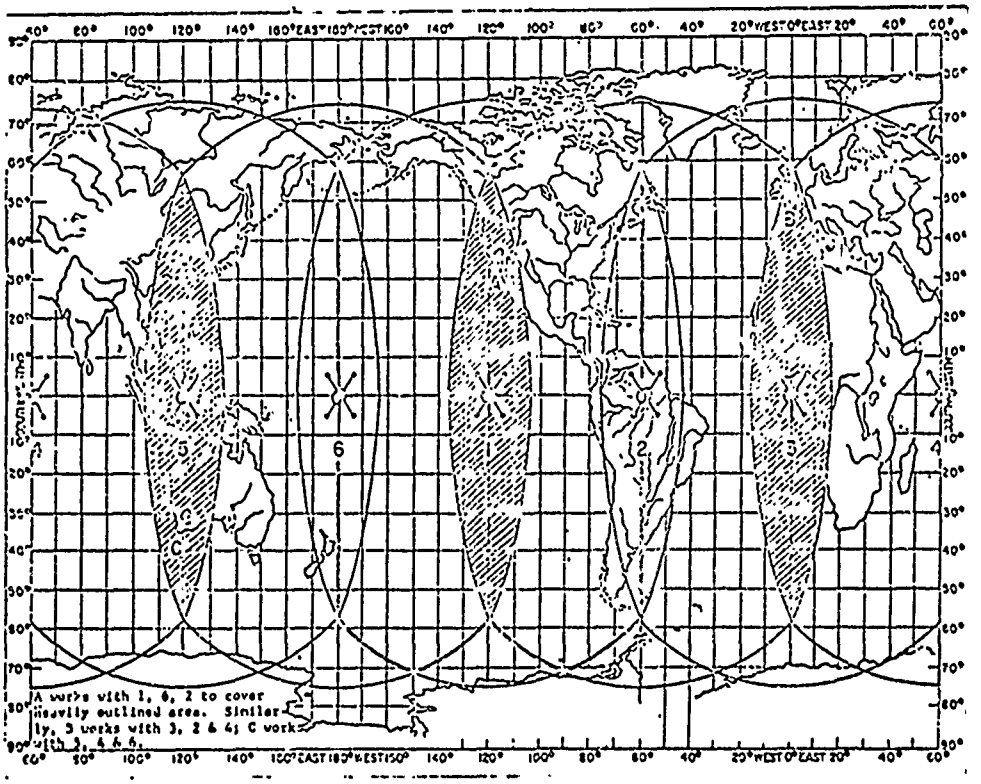


FIGURE 3 COVERAGE BY SIX SYNCHRONOUS SATELLITES, 3 GROUND STATIONS. GROUND STATIONS CAN BE ANYWHERE IN SHADED AREAS.

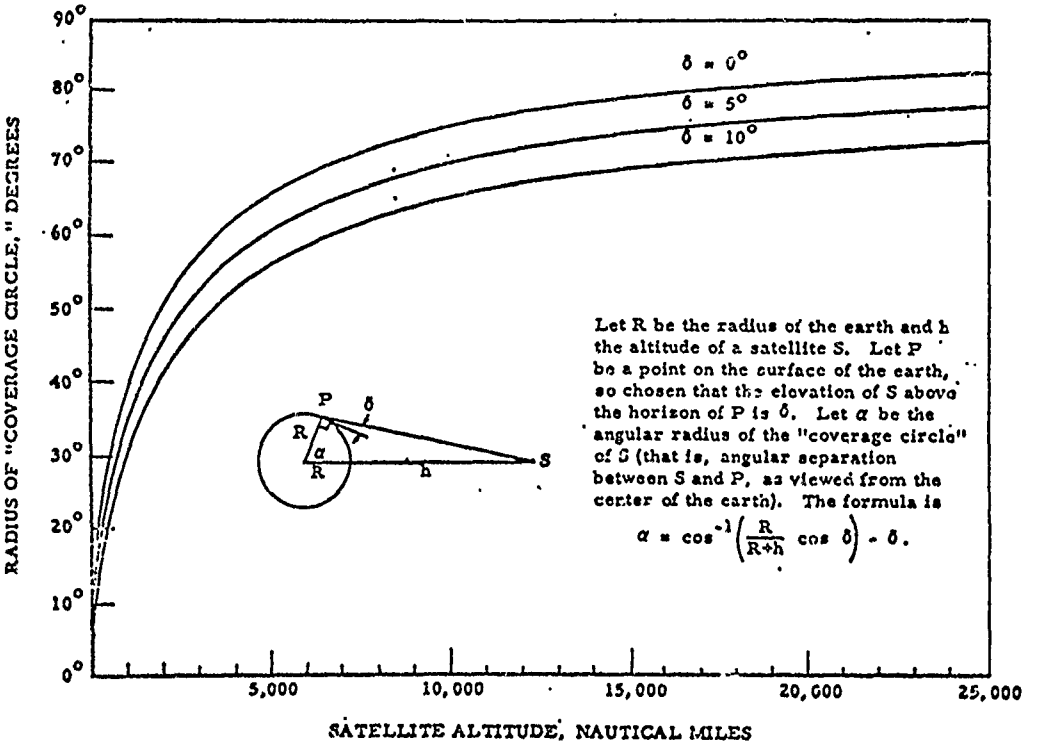
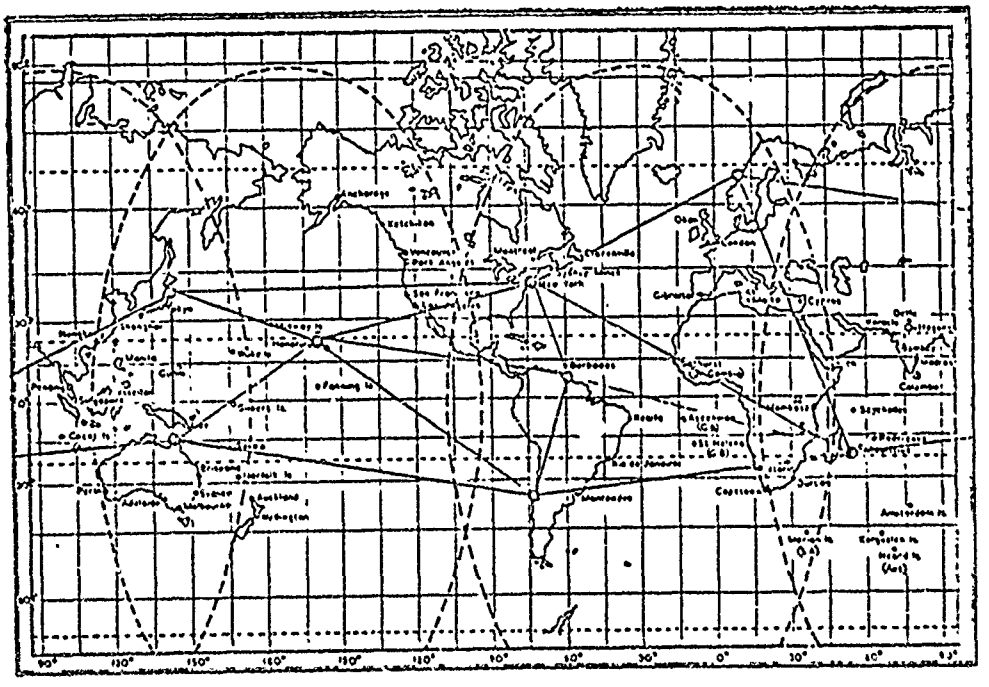


FIGURE 4 SATELLITE "COVERAGE" VS SATELLITE ALTITUDE



NOTE: PROBABLE OMEGA STATIONS LOCATIONS AND SOME OF THE BASELINES ARE ILLUSTRATED.

FIGURE 5 APPROXIMATE LOCUS OF POINTS OF 5° ALTITUDE OF THREE SYNCHRONOUS STATELLITES ON THE EQUATOR AT 30°W 170°W AND 96°E

APPLICATION OF CRYSTAL CLOCKS FOR NAVIGATION AND TIME-ORDERED COMMUNICATIONS

Richard J. Kulpinski
The MITRE Corporation
Bedford, Massachusetts

Summary

Capabilities for position location and real time distribution of information are essential for both civil air traffic and military command and control systems. Time-ordered reporting is being considered seriously for distribution of information. In the techniques and applications being pursued by the MITRE Corporation and presented here, participants maintain an onboard crystal clock synchronized with timing of the reporting net. Transmissions occur automatically based on the onboard clock.

Presented is a technique to achieve synchronization and determine position. Time-of-arrival observations are made on signals emitted by three or more synchronized and located sources of signals. The predictable behavior of the error function of crystal clocks is used to enable a more accurate solution with time of position and synchronization than for the initial computation. The improvement in position accuracy is accomplished without integration of a sequence of position estimates.

Requirements for the crystal and expected accuracy of the solution of position and clock synchronization is presented.

Introduction

Fairly well known is the fact that radio navigation systems involving time-of-arrival measurements are also sources of time and frequency. The discrepancy between the local clock and the system clock, called time offset, is a variable along with position in the analytical description of the physical situation. Solution of time offset is available simultaneously with each solution of position. Because of noise in the measurement process, uncertainty exists in the solution. A common method used to reduce uncertainty due to random noise is integration. Reduction in uncertainty depends on the length of time over which the behavior of the process being integrated is known functionally. For the clock, the behavior of time offset can be described functionally over a long interval of time. Thus, an appreciable reduction in the uncertainty of time offset and therefore clock synchronism is possible. The reduction in uncertainty of position from integration varies depending upon how the aircraft is equipped. With onboard motion sensing and integrating instruments the behavior of the difference between position indicated by the instruments and position obtained from time-of-arrival measurements depends on the error characteristics of the instruments and can be described functionally over a long interval of time. However, in the absence of such instruments, integration is based on the flight path of the vehicle. With frequent maneuvers, the integration interval for position is short and thus reduction in uncertainty of the vehicle's position is small.

This task was performed under Contract XN/ESD; F19628-71-C-0002 with the Electronics Systems Division of the Air Force Systems Command.

The purpose of this paper is to present the concept that the integrated estimate of the value of time offset can be used to enhance appreciably the repeatable accuracy of position even though integration of position is not done.

Application

For the past few years, the MITRE Corporation has been pursuing application and technology of time-ordered reporting for command and control. From analyses of operational problems, it is apparent that periodic broadcasting by each unit of its current position and status is a potential part of the system solution to problems of both military and civilian air traffic operations.

Involved are ground units and aircraft which communicate on a time-ordered net. Once each frame each unit reports its status-position, velocity, altitude, identity, and other sundry information - with a digitally formatted message in an assigned message slot. See Figure 1. Transmissions are broadcast; information in each message is available simultaneously to ground control and airborne units.

Each unit has an onboard crystal clock which it maintains in synchronism with net timing. Each participant's transmission is triggered automatically when the onboard clock indicates that its assigned message slot has arrived.

The requirement is for a means to determine position and synchronize onboard clocks.

Determination of position and synchronization of a local clock is done by receiving signals of three or more established sources with a known or determinable time delay between transmissions. Existing systems such as OMEGA and LORAN are applicable. The MITRE Corporation has also configured the situation in which three or more located and synchronized participants of the time ordered communications net are used as sources for position and timing.

A configuration composed of three or more sources of ranging signals is appealing for it enables user units to become synchronized without requiring transmissions by these units. This allows the onboard clock to be used to support accurate navigation for any receiver equipped user even though the user is not an active participant of the communications system.

The application of an onboard clock to support time ordered transmissions is evident. The application of the clock to support improved navigation is not as visible. In fact, it appears often that an onboard synchronized clock is considered a burden if only position determination is required. For example, time difference techniques are lauded as enabling position determination without the requirement for an onboard synchronized clock. Justification for a clock is often solely for timing applications divorced from the navigation problem. However, timing is very much a part of navigation. In fact, for a moving vehicle an onboard clock, synchronized from the navigation system, enables a significantly more accurate solution

of position than obtained using just the time difference technique. This application alone may be ample justification for an onboard clock.

In the applications being pursued by The MITRE Corporation, the onboard clock supports timing for transmissions in the time ordered net and improved accuracy of position. In the specific configuration mentioned previously in which the signals of the time ordered net are used for ranging and dissemination of time, signals propagate line-of-sight. Airborne units with accurate location and time can be used to extend the positioning capability and net timing to units beyond line-of-sight of ground references. This is possible because of the synegistic value of the clock to both timing and position determination. The capability to extend coverage enables two or more areas to have the same time. Using the time ordered net, transmissions that are broadcast by units in coverage of both areas can be received by both without interference and with a single transmission.

Initial Synchronization and Solution of Position

A unit accomplishes initial synchronization or update by noting the time-of-arrival of signals from three or more locations relative to its local clock. Figure 2 is a timing diagram and shows two axes: one for sources and one for the receiver. The top axis implies the transmission characteristic of general purpose navigation systems--that signals are transmitted with known periodicity and delay between transmissions. The second axis indicates that the onboard clock has periodicity of some specified relationship to the transmissions of the navigation system. The clock possibly generates or times a replica of the transmission pattern. Initially the clock is likely to be out of step with the transmissions of the navigation system. The discrepancy between the user's clock and system timing is called time offset. The value of this parameter is estimated enabling the user to transmit on schedule. The observation made on each signal is the amount of time that elapses between a known division of the local clock and the receipt of the signal. When multiplied by the velocity of propagation this becomes a measured range It differs from the actual range because of the time offset of the local clock. The relationship among the terms time offset, observed range and true range is shown in Figure 3 and is often called pseudo ranging.

There are three unknowns. Given three observations, a sufficient set of simultaneous equations exist:

$$\begin{aligned} \theta_1(m) &= \sqrt{(X-X_1)^2+(Y-Y_1)^2} - \Delta R \\ \theta_2(m) &= \sqrt{(X-X_2)^2+(Y-Y_2)^2} - \Delta R \\ \theta_3(m) &= \sqrt{(X-X_3)^2+(Y-Y_3)^2} - \Delta R. \end{aligned} \tag{1}$$

Figure 4 shows a graphical solution of position and time offset. Each dashed arc has a radius equal to the observed range. Initially, the three arcs do not have a common intersection. The value of time offset is the common increment added to each radius which results in three arcs with a common intersection. The coordinates of the intersection is the user's location.

A linearized form of equations (1), suitable for computer solution, is presented in Appendix I. Although only three equations are given, redundancy provided by more than three sites can be handled using a least squares approach and is also described in Appendix I.

It is interesting to note that the locus of points of the intersection of pairs of arcs of Figure 4 form hyperbolas of constant difference between pairs of observed propagation times. Pseudo ranging, equation 1, is related to the familiar time difference equations:

$$\begin{aligned} \theta_1(m)-\theta_2(m) &= \sqrt{(X-X_1)^2+(Y-Y_1)^2} - \sqrt{(X-X_2)^2+(Y-Y_2)^2} \\ \theta_2(m)-\theta_3(m) &= \sqrt{(X-X_2)^2+(Y-Y_2)^2} - \sqrt{(X-X_3)^2+(Y-Y_3)^2}. \end{aligned} \tag{2}$$

One way of solving for n equations with n unknowns is by successive elimination of variables. When this approach is applied to the variable ΔR in equation (1), time difference indicated by equation (2) results.

Because of receiver noise, and systematic errors, there is uncertainty in the observation and thus uncertainty in the solution. Because both of the above sets of solution equations are equivalent, the solution accuracy of both is the same. The resulting variance in the error of the solutions normalized to time-of-arrival uncertainty is shown in Figure 5 for position, and in Figure 6 for time offset. The accuracy shown in Figure 5 applies also for time difference techniques. The initial solution is sufficiently accurate so that transmissions can be made in the time ordered net without causing interference.

If time offset were known perfectly then the below set of equations, called direct ranging would be appropriate:

$$\begin{aligned} \theta_1(m)+\tilde{\Delta R} &= \sqrt{(X-X_1)^2+(Y-Y_1)^2} \\ \theta_2(m)+\tilde{\Delta R} &= \sqrt{(X-X_2)^2+(Y-Y_2)^2} \\ \theta_3(m)+\tilde{\Delta R} &= \sqrt{(X-X_3)^2+(Y-Y_3)^2}. \end{aligned} \tag{3}$$

The known value of time offset is denoted $\tilde{\Delta R}$. It is a characteristic of the direct ranging approach that the variance in the solution of position due to random uncertainty in the observation is less than for the solution where ΔR is an unknown. However to apply the direct ranging approach an accurate estimate of ΔR must be known.

It is possible to obtain an estimate of ΔR with small variance by solving for time offset at periodic intervals using pseudo ranging equation (1) and then reduce the uncertainty in the solution error by integration over a number of solution periods. The following presents this notion.

Integration as a Means of Reducing Solution Uncertainty

Integration or tracking is a common means of reducing uncertainty. Conceptually, there exists a sequence of single frame solutions,

$$\begin{aligned} X_1 \quad X_2 \quad \cdots \quad X_n \quad \cdots &\rightarrow \boxed{\text{Filter}} \cdots \tilde{X}_n \\ Y_1 \quad Y_2 \quad \cdots \quad Y_n \quad \cdots &\rightarrow \boxed{\text{Filter}} \cdots \tilde{Y}_n \\ \Delta R_1 \quad \Delta R_2 \quad \cdots \quad \Delta R_n \quad \cdots &\rightarrow \boxed{\text{Filter}} \cdots \tilde{\Delta R}_n \end{aligned}$$

Each sequence can be filtered with an integration interval that is as long as the process represented by the sequence can be described functionally.

With stationary receivers a good deal of integration for position is possible. This is also true for aircraft with onboard motion sensing and integrating instruments. The behavior of the difference between position indicated by the instrument and position obtained from time of arrival measurements can be described functionally over a long interval of time. However, due to cost and complexity, not all aircraft carry such instruments. In the absence of these instruments, integration of position is based on the flight path of the vehicle. For aircraft, maneuvers are frequent. Functionally describing flight motion more complex than constant velocity is not practical. Because constant velocity motion usually applies only for short intervals, reduction in uncertainty of position is likely to be small.

For the clock, a function can be written which describes the behavior of the sequence of time offset values over a long interval of time. Furthermore, clocks are available which have an acceptably low sensitivity to accelerations caused by maneuver or vibration. Thus, the error function of the clock can be treated as essentially invariant to maneuvers. A large sample size of single frame estimates of time offset can be used to reduce significantly the uncertainty in the estimated properties of the clock.

Using a Priori Knowledge of Clock to Reduce Uncertainty of Solution of Position

It is generally known that if the clock were known a priori to be perfectly synchronized, a more accurate solution of position would result by using direct ranging, equations (3). Although a perfectly synchronized clock is not a reality, the filtered estimate of the properties of the clock can be used such that the accuracy of estimated position improves each frame with accuracy in the knowledge of time offset even though integration of position estimates over time is not done.

Figure 7 is a block diagram of a possible implementation. Each frame a solution of time offset is obtained using just the observations of the frame. This estimate is an input along with prior estimates to a filter with a long integration interval. The output of the filter, predicted to the time of the next frame, is treated as a priori data along with observations of the next frame in the solution of position. One possible algorithm is to substitute the predicted value of time offset from the filter output for the variable time offset in the ranging equation. A solution of the remaining variables X, Y is computed as shown in the Appendix.

Performance

If observations and update occur at a rate of once per second, then over a 10 minute period an appreciable sample size of 600 would be available for filtering time offset. The ratio of the variance of the output of the filter to the input is about 4/n when filtering a sequence with a constant first derivative such as would exist when using a crystal clock. Therefore, by filtering, an improvement in standard deviation of time offset of at least 10:1 is possible.

Figure 8 shows performance for three circumstances, (1) initial solution of time and position using just one observation from each site; (2) solution of position using filtered estimates of clock offset, and (3) solution of position given a perfectly synchronized clock. A modest sample size of 30 is assumed for filtering estimates of time offset for these curves.

The improvement in repeatable accuracy of position for situation 2 relative to situation 1 increases as the uncertainty of position that is computed without use of filtered time offset increases. This improvement of accuracy is applicable during and following a maneuver. To obtain the same improvement by integration of estimates of position would require a sample size about four times the improvement in variance. However to validly apply such a sample size requires that constant velocity motion be maintained during the time it takes to obtain the sample size.

The dashed lines of Figure 8 show performance when using an atomic standard and integration with a sample of 30. In such a situation the clock's oscillator is assumed to have constant frequency over the integration period, and the filter is synthesized accordingly. Because of this, reduction of the standard deviation of time offset is proportional to 1/n. The accuracy of position for fixed sample size is better with an atomic than a crystal standard. However, the accuracy for both will asymptotically approach that with a perfect clock as the sample size is increased. Sample sizes of 300 and larger are easily justified and result in a negligible difference in accuracy between implementations using atomic and crystal oscillators.

Systematic errors not treated as solution variables cause systematic errors in the solution relative to an absolute reference. These are not eliminated or reduced by integration. However, for control systems, repeatable or relative accuracy is important. Bias or systematic errors in the solution are not of consequence if they are the same for coordinating units of the control system.

Because equations (1) and (2) are equivalent, the same systematic error results in the solution for both the pseudo ranging and time difference approaches. The same error is experienced also in the direct ranging solution, equations (3) if the estimate of time offset, ΔR, is obtained from equations (1) and to the extent that measurement or modeling errors are constant throughout the integration interval. This is a comfortable consequence for users of time difference and the direct ranging approaches will both be in the same relative grid.

Model of Clock Error Function

Attention is given to four items, frequency offset, long-term stability, acceleration sensitivity and short-term stability.

Error in frequency translates into an error in measured range. Crystals are readily available with frequency offset within one part in 10^6 of the nominal frequency. Such an error will result in an error of 2 feet in measured range at 300 nm.

In the algorithm of the filter, time offset is modeled as changing at a constant rate. This is an approximation of the actual behavior. The fact that the frequency of crystal oscillators changes at a linear rate places a quadratic term in the error function of time offset. If the quadratic term is dropped in the filter equations, a reduction in computation time results. With such a simplification, a systematic error of less than four feet results for an integration interval of ten minutes, if the long-term stability is less than 1 part in 10^8 per day. Crystals are readily available which meet this stability after power is applied for a few hours.

Acceleration of maneuvers cause the crystal's frequency to shift. With an acceleration sensitivity of better than 1 part in 10^{10}/g, a 2g maneuver for 30 seconds results in an error in range of less than three feet. Such sensitivity is necessary because errors as the result of maneuvers within an integration interval tend to be cumulative.

Short-term stability defines random uncertainty of oscillator's phase. It implies a limit to useful integration. We would be certainly fooling ourselves if we

suggested accuracy from integration better than the inherent uncertainty in the devices themselves. Short term stability of currently available crystal devices does not force, in practice, a limit on integration.

Summary

An onboard crystal clock can be accurately synchronized to the timing of the navigation system using signals emitted by the navigation system. For moving vehicles in which fruitful integration of position is not likely, the onboard clock enables a marked improvement in repeatable accuracy of position.

The notions discussed here have been simulated. They are now in the process of being exercised using actual ground and airborne units in the Hanscom Testbed, Bedford, Massachusetts.

Appendix I

Consider the situation in which transmissions from p sites occur essentially simultaneously each frame.* Then for the n^{th} frame

$$\begin{matrix}\theta_1(m) \\ \vdots \\ \theta_p(m)\end{matrix} = \begin{matrix}\sqrt{(X_n-X_1)^2+(Y_n-Y_1)^2+\Delta Z_1^2} - \Delta R_n + N_1 \\ \vdots \\ \sqrt{(X_n-X_p)^2+(Y_n-Y_p)^2+\Delta Z_p^2} - \Delta R_n + N_p\end{matrix} \qquad \text{(A-1)}$$

is a set of simultaneous equations; where

$\theta_i(m)$ = measured range made on signal of i^{th} site,

X_iY_i = location of i^{th} source,

ΔZ_i = difference in altitude between source and receiver taken from altimeter data,

X_n, Y_n, ΔR_n = solution variables for receiver's position and time offset,

N_1 - - - N_p = random error in time-of-arrival measurement, mean zero, and variance σ_m^2.

Linearization of equations (A-1) about an assumed solution X_a, Y_a, ΔR_a using a Taylor series expansion enables a solution using a digital machine. For the i^{th} row of equation (A-1) then,

$$\theta_i(m) = \theta_i(a) + \frac{X_a-X_i}{R_i(a)}(X-X_a) + \frac{Y_a-Y_i}{R_i(a)}(Y-Y_a) - (\Delta R-\Delta R_a)$$

where

$$\theta_i(a) = R_i(a) - \Delta R_a,$$

$$R_i(a) = \sqrt{(X_a-X_i)^2 + (Y_a-Y_i)^2 + \Delta Z_i^2}, \qquad \text{(A-2)}$$

X_a, Y_a, ΔR_a = Best estimate of solution for X, Y, ΔR.

Define

$$B_{ix} = (X_a-X_i)/R_i(a)$$

$$B_{iy} = (Y_a-Y_i)/R_i(a)$$

$$D\theta_i = \theta_i(m)-\theta_i(a).$$

* This development is easily extended for the situation in which transmissions of a group occur separated in time.

Then in matrix notation equation (A-1) after linearization is

$$\begin{bmatrix} D\theta_1 \\ \vdots \\ D\theta_p \end{bmatrix} = \begin{bmatrix} \beta_{1x} & \beta_{1y} & -1 \\ \vdots & \vdots & \vdots \\ \beta_{px} & \beta_{py} & -1 \end{bmatrix} \cdot \begin{bmatrix} (X-X_a) \\ (Y-Y_a) \\ (\Delta R-\Delta R_a) \end{bmatrix}. \qquad \text{(A-3)}$$

Define equation (A-3) as

$$D\theta = J \cdot DS.$$

If only the observations made during a frame are used as data for the solution then a least squares solution is

$$\begin{bmatrix} X_n \\ Y_n \\ \Delta R_n \end{bmatrix} = \begin{bmatrix} X_a \\ Y_a \\ \Delta R_a \end{bmatrix} + (J^TJ)^{-1} J^T D\theta. \qquad \text{(A-4)}$$

In the above, each observation is given equal weight.

In the initial frame, X_a, Y_a, ΔR_a is arbitrary but close to the actual value. This may not be sufficiently close to the actual solution to justify absence of the second and higher order terms of the Taylor expansion. In such a case, the algorithm is iterated; the solution is used as the value of X_a, Y_a, ΔR_a and equations (A-2) to (A-4) computed again. Iteration is stopped when the incremental change in the solution is small. In succeeding frames, X_a, Y_a, ΔR_a is taken as the prior solution predicted to the time of the current observations.

Denote $\widetilde{\Delta R}_n$ as the filtered value of the sequence of single frame estimates of time offset. An algorithm which uses the filtered estimate of time offset along with observations of the next frame to give a better solution accuracy for position is as follows. Linearization of equation (3) gives

$$\begin{bmatrix} D\theta_1 + \Delta R_n \\ \vdots \\ D\theta_p + \Delta R_n \end{bmatrix} = \begin{bmatrix} \beta_{1x} & \beta_{1y} \\ \vdots & \vdots \\ \beta_{px} & \beta_{py} \end{bmatrix} \cdot \begin{bmatrix} (X_n-X_a) \\ (Y_n-Y_a) \end{bmatrix}. \qquad \text{(A-5)}$$

Define equation (A-5) as

$$DR = A \cdot DP.$$

Then, a solution is obtained by computing

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} = \begin{bmatrix} X_a \\ Y_a \end{bmatrix} + (A^TA)^{-1} A^T DR. \qquad \text{(A-6)}$$

References

1. "Study of Methods for Synchronizing Remotely Located Clocks, National Aeronautics and Space Administration, NASA CR-738, March 1967.
2. L. Horowitz, "Direct-Ranging LORAN," Navigation, Journal of the Institute of Navigation, Vol. 17, No. 2, Summer 1970, pp. 200-204.
3. E. A. Gerber, R. A. Sykes, "State of the Art - Quartz Crystal Units and Oscillators," Proceedings of the IEEE, Vol. 54, No. 2, February 1966, pp. 103-116.

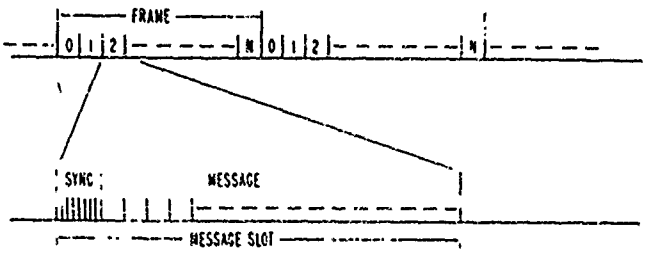


FIGURE 1 TIMING-TIME ORDERED NET

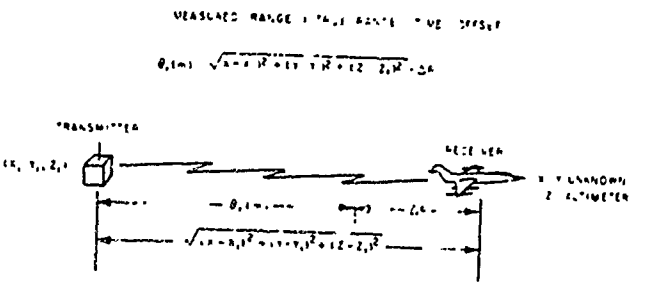


Figure 3 RANGING EQUATION

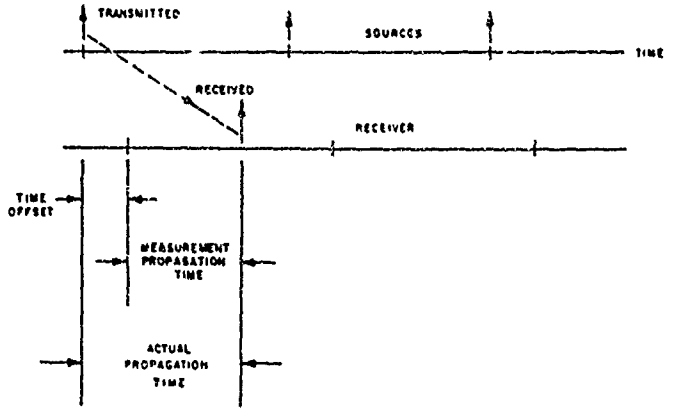


Figure 2 RANGING MEASUREMENT

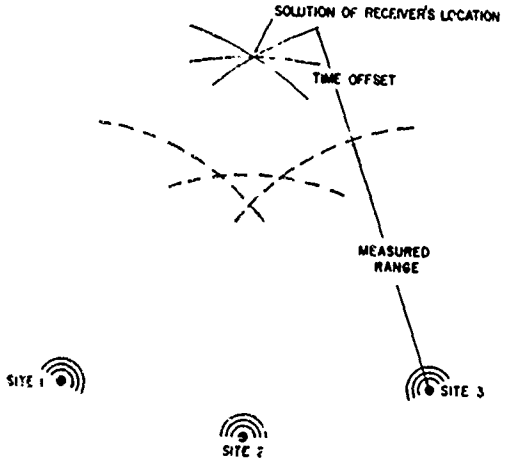


Figure 4 GRAPHICAL SOLUTION OF TIME OFFSET AND POSITION

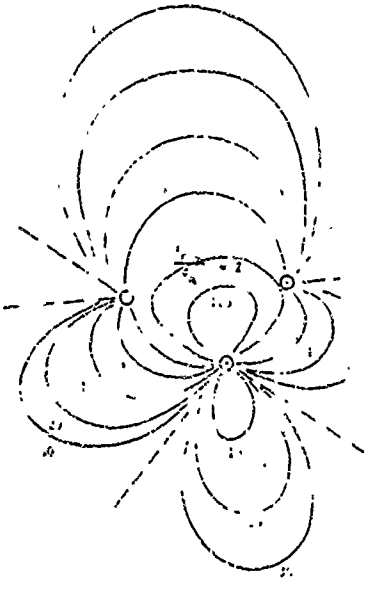


FIGURE 5 POSITION-CONTOURS OF CONSTANT UNCERTAINTY, PROB (RADIAL ERROR $< d_{rms}/\sigma_m$) ≈ 63

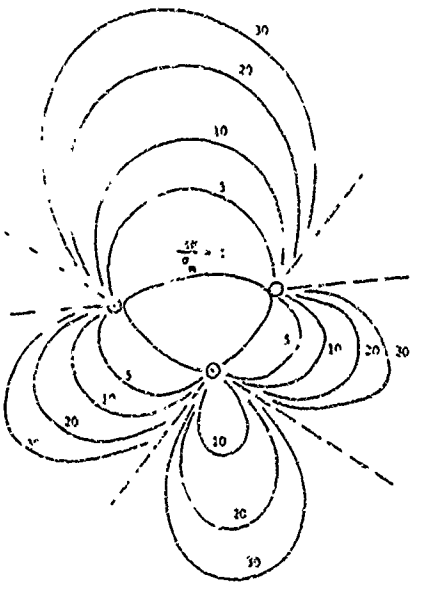


FIGURE 6 TIME OFFSET-CONTOURS OF CONSTANT STANDARD DEVIATION

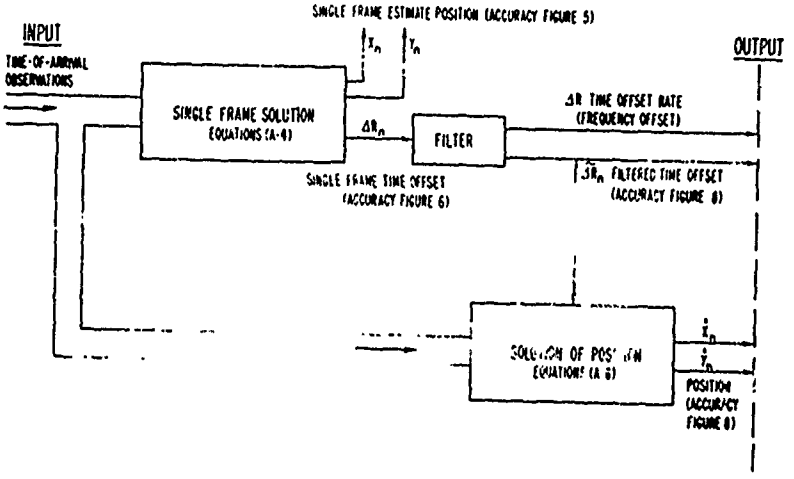


FIGURE 7 BLOCK DIAGRAM IMPLEMENTATION AND SEQUENCE OF EVENTS

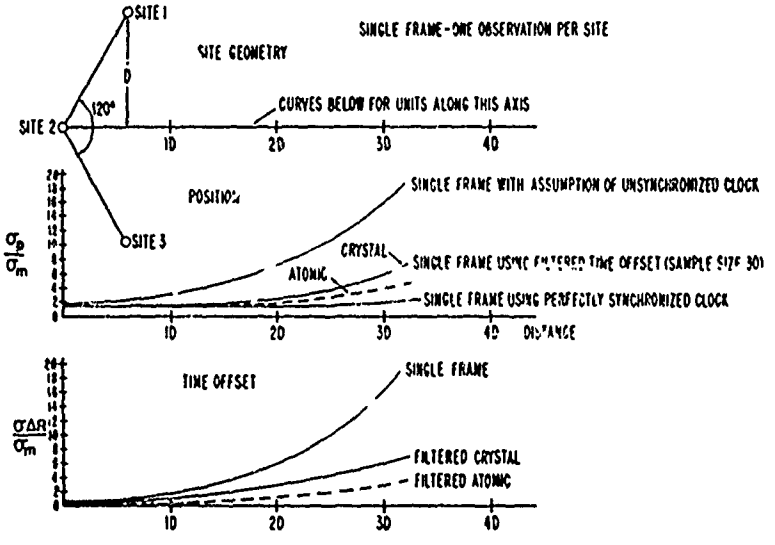


FIGURE 8 COMPARISON OF PERFORMANCE

TIME SYNCHRONIZED RANGING SYSTEM

(TSRS)

ROBERT M. AUGHEY
SINGER-GENERAL PRECISION, INC.
KEARFOTT DIVISION
LITTLE FALLS, NEW JERSEY

Summary

Singer-General Precision, Kearfott Division, has developed and is currently field testing a time-ordered precision ranging and relative position location system identified as TSRS. The TSRS is based on cooperative synchronization of both time and frequency between up to 1024 LOS users to a relative time accuracy of less than 7 nanoseconds. Additional TSRS capabilities include automatic user identification and two-way communication data link. The TSRS is currently under field ground test and will be flight tested under a U.S. Navy Contract in 1971.

The Kearfott Division of Singer-General Precision has, under Company sponsorship, developed a cooperative ranging system employing local frequency sources and clocks which are precisely synchronized in relative frequency and time to a designated control station within a community.

Kearfott identifies this system as (TSRS) a Time Synchronized Ranging System.

A general definition of this type of equipment is a collection of units, or subscribers, located remotely from each other, which employ a common relative time and frequency reference for the purpose of sequencing events, timing events, and determining slant range between units.

Important system by-products of the TSRS process include two-way data communication, unit identification, and relative position location, where adequate ranges are available to perform trilateration computations.

TSRS is a time-ordered system (See Figure 1) in which a system cycle is divided into "N" slots, each of which is precisely 2.62144 milliseconds long. For the purpose of this presentation the 00 or first slot in the cycle will always be assigned to the Control Station, although Kearfott does incorporate a capability for rotating Control to qualified users in the community in order to service users not in line-of-sight of the designated Control Station.

The Control Station always transmits a "Coarse Sync" message in slot 00, initiated at his own T_o. Each user, having been previously assigned an exclusive time slot, requests and receives Fine Synchronization from the Control Station within his own slot.

A simplified diagram of the time format for the synchronization process is shown in (Figure 2). For ease of demonstrating the relative time relationships this illustration has divided a slot into 10 equal units.

The Control Station transmits a Coarse Sync message in slot 00 at his indicated T_o. This message is received by all users within LOS, each of which resets his local clock to T_o on receipt of the message. Obviously, each user clock is now "late" with respect to the Control Station clock by the individual propagation time from the Control Station to each user, however, it is important to note that the Coarse Sync transmission always establishes relative synchronization between User and Control to 600 microseconds or less, corresponding to the approximate propagation delay at 100 miles maximum LOS distance between Control and User.

For illustration, assume that the User of Figure 2 is two time units in distance from the Control Station and that he has been assigned slot 01, immediately following the 00 slot. The User has received the C.S. message and reset his clock to T_o on receipt. 10 units (2.62144 ms.) later, at the start of his assigned 01 slot, the user initiates a Fine Sync Request message to the Control System. Assuming no relative movement between the stations, and therefore two time units for propagation, the F.S. Request arrives at the Control Station 4 units late with respect to the Control clock (two propagation times). The Control Station assumes that the F.S. Request delay of 4 units is all propagation delay (range), and initiates a Fine Sync Reply Message to the 01 User 4 time units ahead of the end of slot 01 which arrives at the User two time units later. The User realizes that the F.S. Reply should have arrived exactly at the end of slot 01 if his time was synchronized with the Control Station, and therefore his clock is "late" by two time units and automatically resets it to cause the start of slot 02 to be coincident with the arrival of the F.S. Reply. The User and Control clocks are now in Fine Synchronization and identical transmissions in the next and successive system cycles can be utilized to measure slant range between units. The actual circuit implementation of the coarse and fine synchronization in TSRS is somewhat more complex than this simplified description and will be discussed in more detail later.

The TSRS message structure has been designed to provide acquisition, time synchronization, data

communication, and automatic unit identification. Figure 3 illustrates the TSRS message structures. All transmissions are preceded by 6 bits of quieting (1.28 μs./bit).

The Control Station's Coarse Sync. transmission in slot 00 consists of two 13 bit inverted Barker codes followed by two 9 bit slot number words and 144 bits available for data. One parity bit for each 8 bits of data is incorporated. This transmission therefore is 248 microseconds maximum in length, and is the only transmission occurring in slot 00.

In slot 01, the User F.S. Request transmission consists of two 13 bit normal Barker codes followed by 144 bits of data, resulting in a message length of 225 μs. maximum. The F.S. Reply message, also in slot 01, consists of two Barker codes, the first inverted and the second normal, resulting in a message length of 41 μs. maximum. The combinations of normal and inverted Barker codes, coupled with slot assignments, serve to identify the User, and the entire community, the type of message and its origin. The Coarse Sync message is transmitted only once per system cycle, but the F.S. Request and Fine Sync Reply messages are transmitted and received by each User in his assigned slot every system cycle.

Figure 4 is a TSRS System Block Diagram. Each TSRS consists of a Receiver-Transmitter Unit (RTU), a Digital Unit (DU), and a Control Display Unit (CDU). The TSRS Control Station is shown with a computer and teletype, however, these are used only to process TSRS range and data for specific applications such as relative position location. Each TSRS unit is fully implemented to perform as a User or Control Station with the choice and slot assignments selected by means of the CDU's.

Figures 5 and 6 are photographs of the TSRS Receiver-Transmitter Unit. The RTU is housed in a 3/4 ATR case and weighs 20 pounds. This unit is an Engineering Model, built in the laboratory, and is not representative of final mechanical design. Figures 7 and 8 are photographs of the TSRS Digital Unit. The DU is housed in a 1/2 ATR case and weighs 15 pounds. This unit is representative of final TSRS packaging and incorporates SGP's X-Y logic cards as well as high speed, multi-layer cards in a mechanical design proven by hundreds of Kearfott production navigational equipments for Military Aircraft.

Figure 9 is a photograph of a TSRS Submaster (Control Station) consisting of the RTU, DU, CDU, Computer, and a ground unit antenna.

A block diagram of the Receiver-Transmitter Unit is shown in Figure 10. The RTU is designed to operate in the 960 to 1215 MHz ("L" Band) range and we are currently field testing at 1140 MHz with a 7 MHz BW and a 1% duty cycle. All RTU functions are controlled by the DU, for example, the transmitter section accepts baseband data at a 780 KHz rate, XMIT key signals, and a 60 MHz input to the bi-phase modulator. The output of the modulator is a 60 MHz carrier whose phase is reversed by the differentially encoded data. This signal is mixed with a 1080 MHz L.O. and the upper side band selected to provide a 1140 MHz bi-phase modulated carrier, which is amplified to a one kilowatt level and delivered to the antenna via a circulator. The transmitter is keyed on just prior to the message and keyed off at the end of the message, while an opposite keying signal is simultaneously applied to the receiver.

An incoming message is passed through the antenna, circulator, and protective limiter to a down converter where the 1140 MHz bi-phase modulated signal is mixed with an 1080 MHz L.O. providing a 60 MHz input to the IF amplifier. This amplifier provides 80db of gain by means of 9 successively cut-off limiter stages and exhibits nearly constant phase shift and group delay throughout its dynamic range. The IF amplifier has a phase response that is essentially insensitive to the limiting level.

The amplified and limited 60 MHz signal is split with one output delayed by one bit (1.28 μs.). The direct output and the delayed output are then compared in the phase detector, providing differential decoding of the 60 MHz bi-phase information. The resultant video signal is applied to a video decoder which supplies time-of-arrival (TOA) pulses, and data to the TSRS Digital Unit.

Figure 11 illustrates the RTU Barker decoder in simplified form. The video output of the phase detector is routed to a 13 bit tapped delay line. The Barker is a special synchronizing code (13 bits for TSRS) which, when cross-correlated with a stored sequence, can provide a single pulse at correlation, the amplitude of which can be set by means of a threshold detector. Both positive and negative thresholds are incorporated for TOA pulses since TSRS uses both normal and inverted Barker codes.

Figure 12 is a block diagram of the TSRS Digital Unit. The Main Timer provides a continuous time base and reference for the unit and when the unit is designated Control its counting length defines the slot interval for the entire system.

The Main Timer operates directly at a 100 MHz rate, derived from the local frequency standard, using ultra high speed emitter-coupled logic, with multi-layer, controlled impedance, printed circuit packaging. All critical timing functions are synchronously clocked, with less critical functions clocked by a ripple carry signal. All of the DU's key timing signals are derived from the Main Timer. The Main Timer also supplies timing control for the Secondary Timer.

The Secondary Timer provides a time base utilized for time measurements and control of all R/T transmissions. The Secondary Timer is periodically "phase-locked" to the Main Timer, however, it can be stopped by the appropriate TOA's to provide a measure of time and/or range.

The Slot Counter identifies each TSRS slot, and are initialized by the CS message from the Control Station during acquisition. Thereafter, it advances one count every time the Main Timer completes its timing of a ranging slot.

The TSRS frequency source is located in the DU. This oscillator-multiplier has output frequencies of 20 MHz and 100 MHz. The unit is voltage controlled, housed in a proportional control oven and has a drift rate of 1×10^{-8} parts per day. Voltage control of this source is obtained from an up-down counter and a precision D/A converter. These elements, coupled with measured time-errors and the Digital Filter, form the oscillator frequency and phase control loop. An up-down counter is initially set to mid-range and incremented or decremented by a series of pulses proportional to the measured synchronization error. The gain associated with a given correction is controlled by the Digital Filter and is adjusted by directing the correction pulse train to various weighted inputs in the up-down counter. The D/A converter closes the frequency and phase control loop by supplying an analog signal to the oscillator as a function of the stored count.

The Digital Filter logic implements an algorithm which smooths time, frequency, and phase corrections. This algorithm is an approximation to a simple Kalman Filter (linear filter with time varying gains). The filter logic accepts a measured time error and processes it in accordance with the filter algorithm to generate a time correction for the Main Timer and frequency or phase corrections for the oscillator. The Main Timer is corrected under control of the filter by inhibiting or adding counts (counting by two's at the normal clock rate) depending on the sign of the error.

The remaining DU functions are associated with TSRS control, data message formatting storage and processing, and general "housekeeping".

Figure 13 is the F.S. state diagram representing the flow of F.S. control in order to accomplish the desired time frequency and phase corrections.

The Digital Filter is active only in the User TSRS, since the Control Station reference is common for the entire system, and no internal corrections are made as long as the TSRS is operating in a "relative" mode, although provisions are incorporated to reference the Control Station (and therefore the system) to any desired external source.

The User DU is initialized by the C.S. process previously described, and for the first update (F.S. Request and Reply) the clock gain (K_C) is one and the oscillator gain (K_O) is zero. For the next four successive updates the clock gain is unity and the oscillator gain is increased. In the following slots the clock gain is reduced by a constant plus a variable depending on the number of missed updates (N) and the oscillator is reduced to low gain. The DU will remain in this state until the measured time error is equal to or less than a preset count. The DU will thereafter remain in the following state unless the measured time error on an update exceeds the preset count. Should this occur, a transfer is made to the adjacent state and if the next update error is equal to or less than the preset count of the previous state, status is returned to the previous state, however, if the next update error is greater than 10 times the preset count, the status is returned to the unity clock and high gain oscillator state and the status proceeds through the state diagram again.

For the case where the update error is greater than the preset count but less than 10 times the preset count, the status is returned to the state where the clock gain is a constant plus a variable based on the number of missed updates. The TSRS Digital Filter was modeled and simulated on an 1108 computer prior to hardware design. Laboratory and field tests of TSRS have verified system performance as predicted by the simulation.

Figure 14 illustrates, in simplified form, the TSRS range measurement signal flow between two cooperative units.

A message is formatted, including the Barker code and data, in the Digital Unit, and, at the proper time, sent to the Receiver-Transmitter Unit as a differentially encoded video signal. This signal is used to bi-phase modulate a 60 MHz carrier which is then up-converted to and amplified at 1140 MHz.

The radiated signal is received by the second unit, down-converted to 60 MHz, amplified and detected with the resultant video sent through the Barker decoder.

This output consists of TOA pulse(s) and serial data which are sent to the Digital Unit, where the TOA's are used to measure range, clock error, etc., and the serial data is buffered for transfer to external equipment.

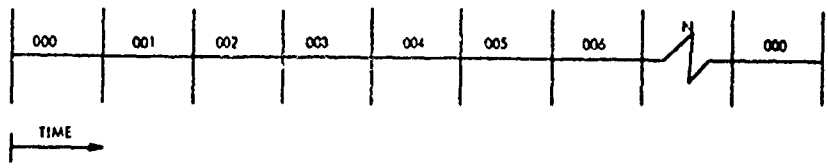
This Transmit-Receive process is reversed at the proper time for message flow in the other direction.

Figure 15 is a summary of the current performance of the TSRS indicating an operational range of 100 NM, provided a line of sight path exists. Slant range measurement accuracy is 21 feet, one sigma, based on a relative time synchronization accuracy of each User to the Control Station of 7 nanoseconds, one sigma. The TSRS provides User identification on a time-ordered assigned slot basis with up to 300 bits of data communication capacity between the Control Station and each User every 2.68 seconds at a data rate of 780 Kbz.

There are many military and commercial applications for TSRS including:

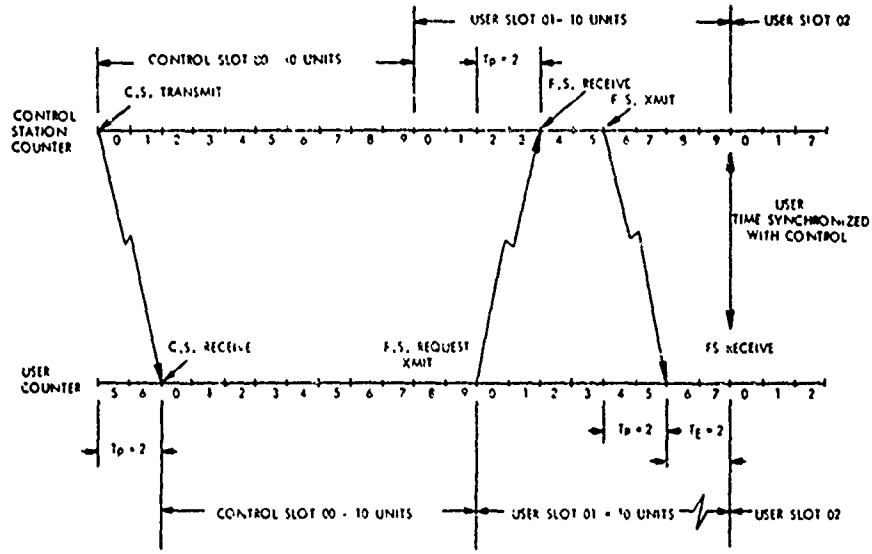
.Relative Position Location
.Navigation
.Precision Reference for Hybrid System
.Command and Control
.Emitter Location Systems
.Collision Avoidance
.Air Traffic Control

TSRS TIME ORDERED FORMAT

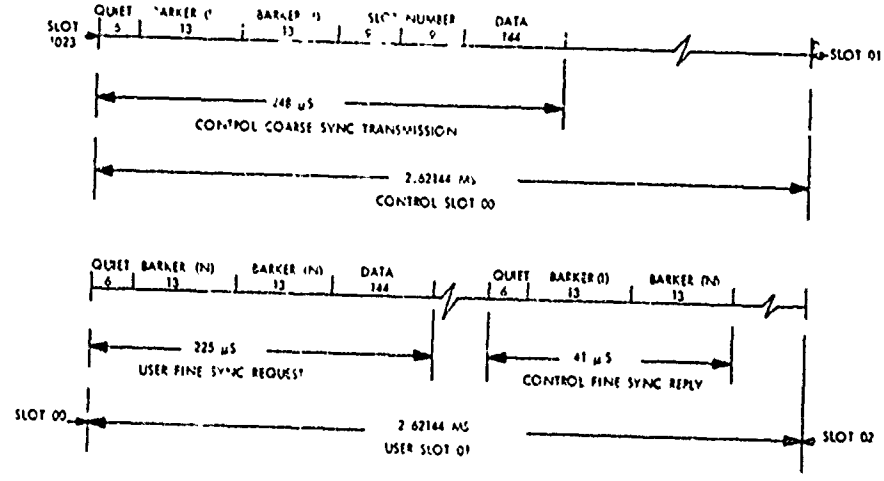


TSRS SYSTEM CYCLE = 1024 SLOTS = 2.68435456 SECONDS
TSRS SLOT = 2.62144 MILLISECONDS

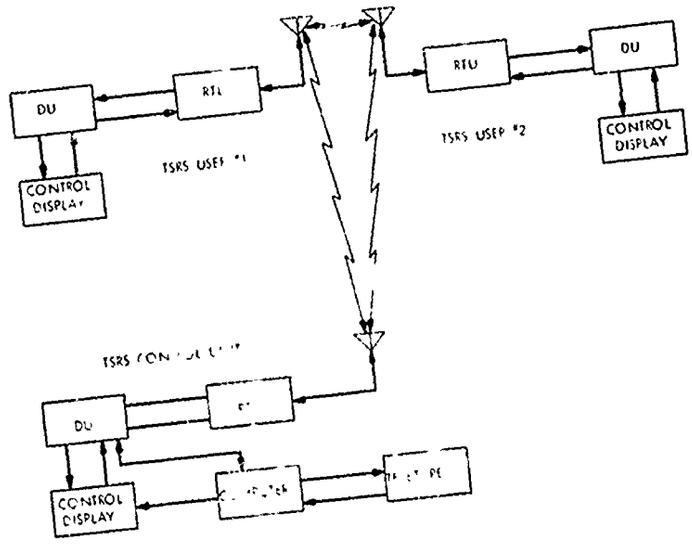
TIME SYNCHRONIZATION FORMAT
TWO WAY RANGING



TSRS MESSAGE STRUCTURE



TSRS SYSTEM BLOCK DIAGRAM



TSRS RECEIVER-TRANSMITTER UNIT

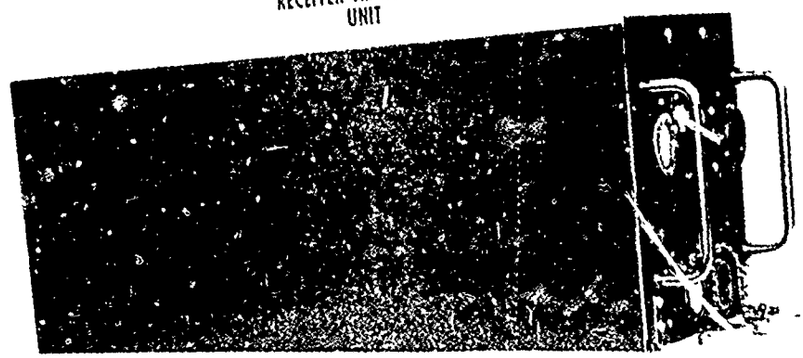


PHOTO NO 33034

SINGER

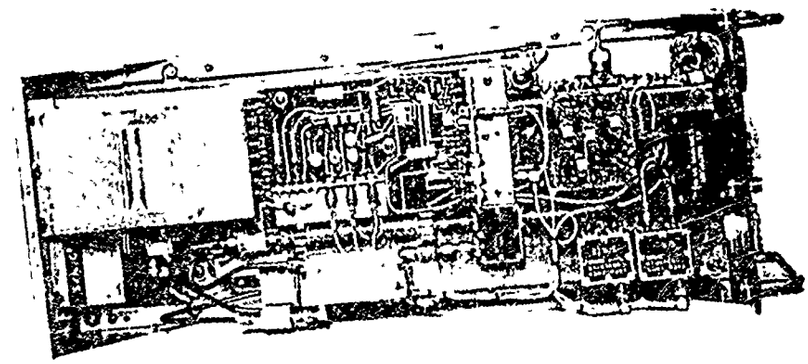
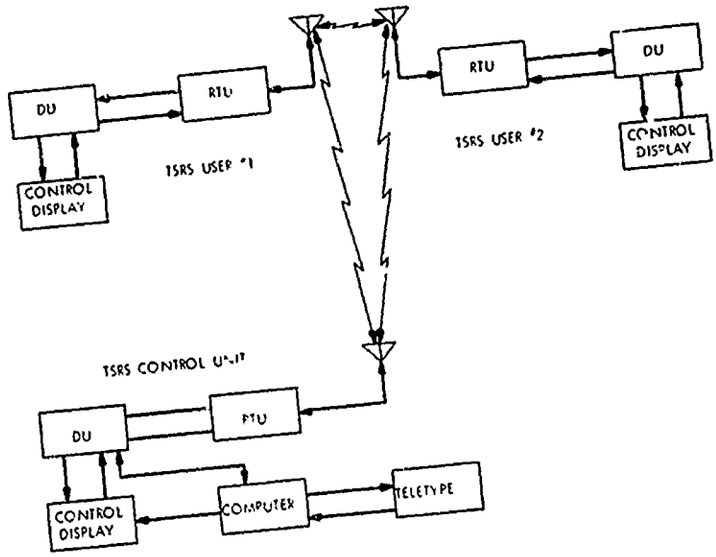


PHOTO NO 33030

SINGER

TSRS SYSTEM BLOCK DIAGRAM



TSRS RECEIVER-TRANSMITTER UNIT

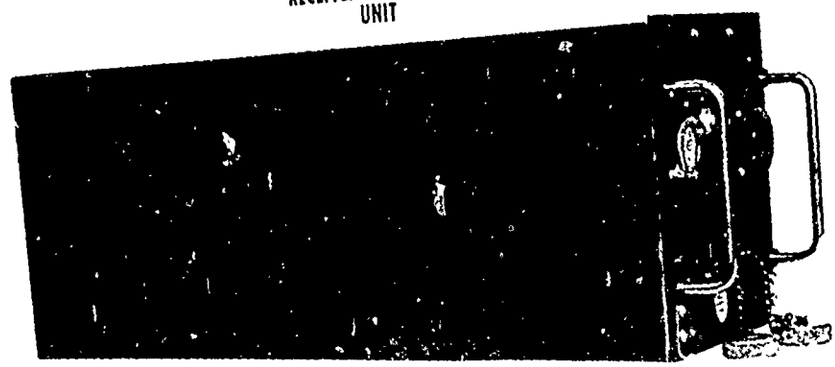


PHOTO NO 33029

SINGER | SINGER-GENERAL PRECISION, INC. LITTLE FALLS, NEW JERSEY

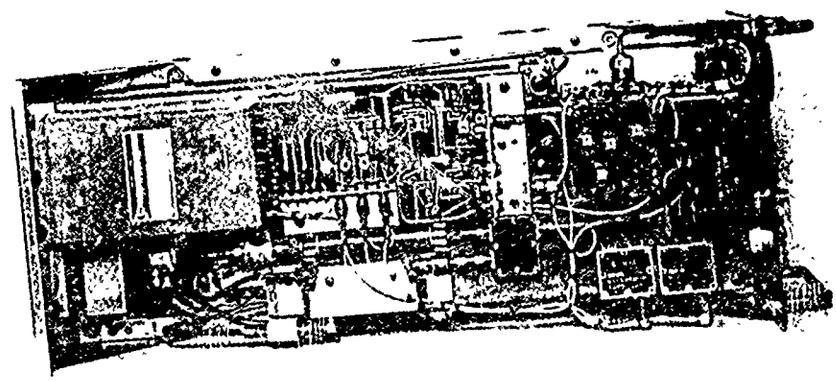
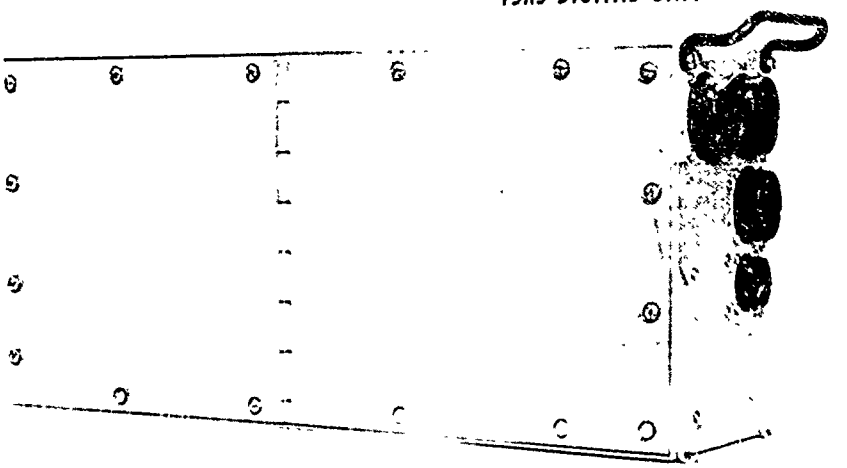


PHOTO NO 33030

SINGER | SINGER-GENERAL PRECISION, INC. LITTLE FALLS, NEW JERSEY

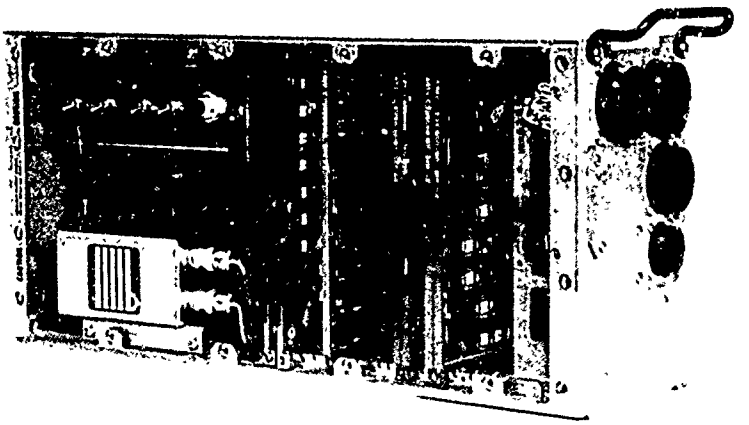
TSRS DIGITAL UNIT

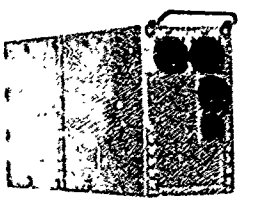


SINGER | SINGER GENERAL PRECISION, INC LITTLE FALLS NEW JERSEY

PHOTO NO 33035

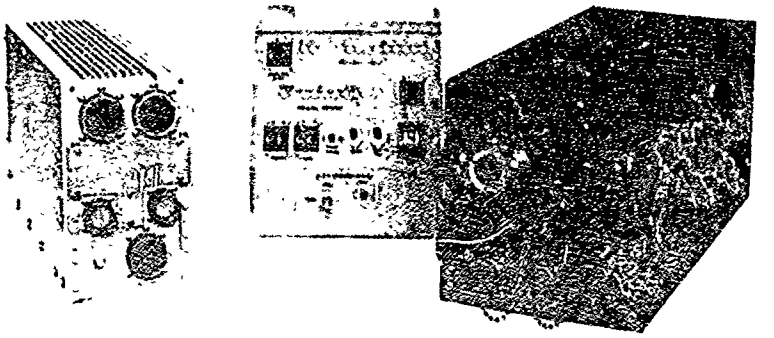
TSRS DIGITAL UNIT





SINGER | SINGER GENERAL PRECISION INC LITTLE FALLS NEW JERSEY

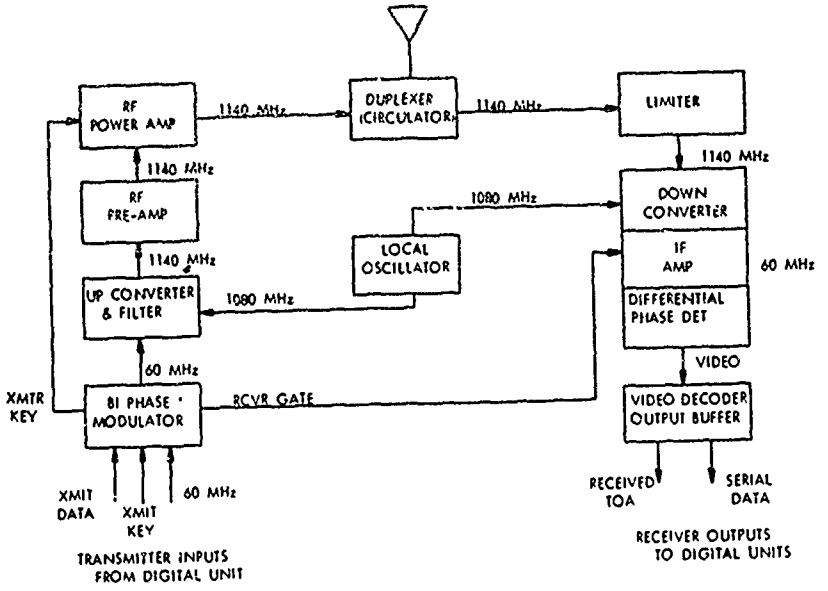
TSRS SUBMASTER



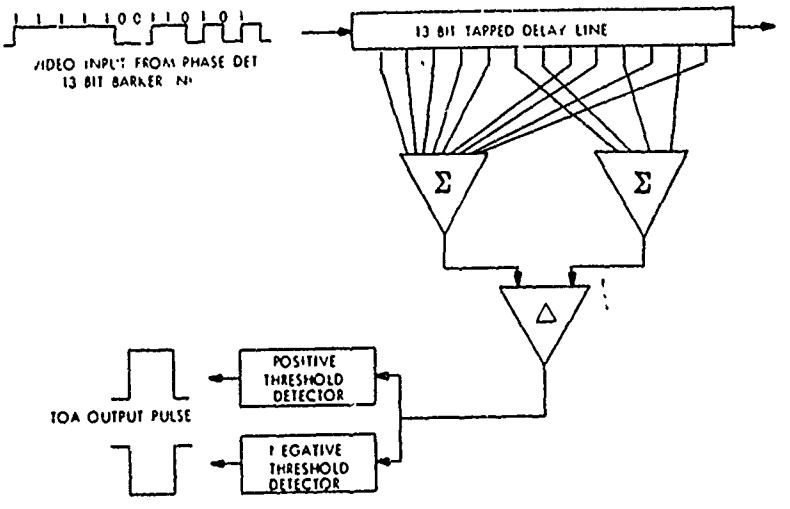
SINGER | SINGER GENERAL PRECISION, INC LITTLE FALLS NEW JERSEY

PHOTO NO 33040

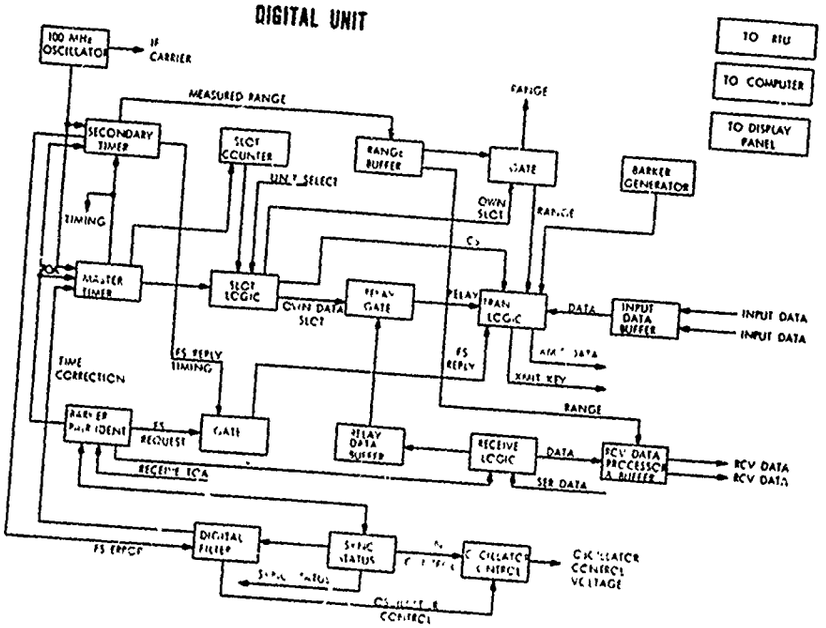
FUNCTIONAL BLOCK DIAGRAM
TSRS RECEIVER-TRANSMITTER UNIT



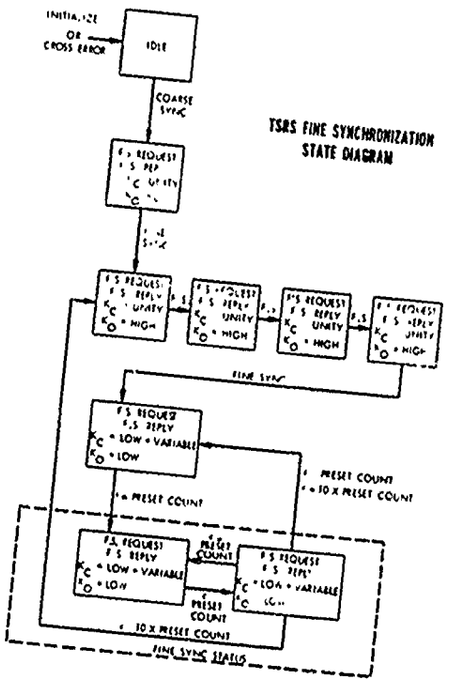
TSRS BARKER DECODER BLOCK DIAGRAM



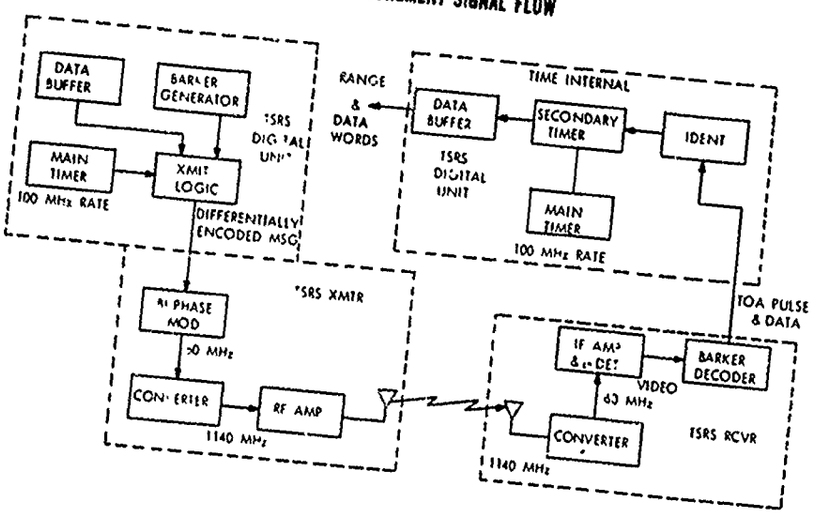
DIGITAL UNIT



TSRS FINE SYNCHRONIZATION STATE DIAGRAM



SIMPLIFIED BLOCK DIAGRAM TSRS RANGE MEASUREMENT SIGNAL FLOW



TSRS PERFORMANCE SUMMARY

• OPERATIONAL RANGE	100 NM - LOS
• SLANT RANGE ACCURACY	21 FEET - 1σ
• RELATIVE SYNCH. ACCURACY	7 NANOSECONDS - 1σ
• COMMUNICATION CAPACITY	300 BITS/SYSTEM CYCLE
• DATA RATE	780 KHz
• SLOT WIDTH	2.62144 MILLISECONDS
• SYSTEM CYCLE	2.68435456 SECONDS
• NUMBER OF USERS	1023 MAX.
• OPERATING FREQUENCY	1140 MHz
• MODULATION	BI-PHASE
• RF BANDWIDTH	7 MHz
• RF POWER	1 KW 1% DUTY CYCLE
• ANTENNAS	OMNI - "L" BAND
• MASTER OSCILLATOR	100 MHz OUTPUT SELF-CONTAINED

"Piezoelectric Sensors for Use as Pollution Detectors, Meterology Monitors and Research Instruments."

J. Kertzman, Oceanport. New Jersey

The major research and dynamic growth of new industries in the '70's will be directed toward our environment. Three of the major areas for environmental activity will be in Air and Auto Pollution, Manufacturing Processes and New Products and Research by the U.S. Government Laboratories, Industry and Universities to meet both requirements.

The five man made major sources of air pollution are automobiles. major industry plants power plants, space heating and refuse disposal The pollutants include carbon monoxide hydrocarbons oxides of nitrogen, sulfur oxides, particulates and a variety of other combustion and processing products.

The industries whose products and processes are the major contributors to pollution are:

Automobile Manufacturers who will have to redesign their power systems or add devices to reduce emissions. They will be responsible for measuring and controlling the auto emissions. The auto manufacturers are spending about 400 million dollars a year on this problem. They must provide information, recommendations and possibly equipment to 60,000 new car dealers who will have to service the vehicles.

The Petroleum Industry will have to develop and produce new fuels to meet the new pollution requirements. New processes are being developed and the refineries will also be required to monitor their plants that produce these new fuels. In addition. the petroleum industry will also be recommending emission measuring equipment to their service stations for measurement and correction procedures in the new Federal Standards. There are 140,000 service stations in the United States.

Petrochemical Chemical, Chemical Processing and other Manufacturing Plants will require modifications in their processes to meet new air pollution regulations. Monitoring apparatus will have to meet approved requirements which will be established by State and Federal Government Agencies.

Electric Utilities, Apartment Houses, Hotels, Office Buildings, Warehouses and Other Large Fuel Consumers and Local and Private Refuse Disposal Stations will have to be monitored to insure compliance with pollution standards.

The inspection and enforcement of the standards will be made by State and Local pollution administrations. Air and auto pollution standards will require continuous moritoring and periodic inspections. The State of New Jersey Environmental Protection Agency is enforcing a state code on air pollution. It is also evaluating an advanced car inspection system which will mean 100% car inspection to meet new Federal Specifications.

An estimate of the air pollution instrumentation market for 1970-1980 was presented at the APCA meeting in June, 1970, by R. Bertram of Esso Research and Engineering Co. at 469 million dollars.

"Is there any simple, reliable, accurate and precise instrument that will meet the requirements of this vast program?" The answer is "no" and time is running out.

There are many prospects on the list, and several instruments based on old wet chemical methods are being used; however, the scope of this is so large that a new type of sensor is required.

In my opinion, a leading candidate for a universal pollution sensor is the coated piezoelectric crystal. It has many outstanding features such as high sensitivity. low cost, simple instrumentation, detection of specific compounds, reversible mass changes. sensors that are mounted on site or remote and the units can be adapted to simple pass-fail indicators or controllers.

The first commercial coated quartz crystal instrument was introduced in 1963 and today is used worldwide as a water vapor analyzer in all of the leading refineries throughout the free world. It is also used in Petrochemical, Chemical, Natural Gas, Atomic Energy and other manufacturing and research facilities. It has a sensitivity of 0.05 ppm (v/v) water vapor. and it has a faster response to dynamic changes in water vapor concentration than any other hygrometer.

Other applications of the coated quartz crystal have been as specific detectors for hydrocarbons. carbon dioxide. sulfur dioxide and particulates.

The linear temperature response and high sensitivity of the LC crystal has been demonstrated in commercial quartz crystal thermometers. Demonstration of the quartz as a meterology detector for humidity in the upper atmosphere has also been reported. Two European countries are using coated quartz crystals as humidity sensors in radiosondes. The coated quartz crystal has a higher sensitivity for measuring barometric pressure than the present bellows used in the Weather Bureau ESSA radiosondes. The present GSA requirements are to 250,000 sondes/year. Overseas requirements will double the potential. The sondes for the weather balloons are disposable so the crystal detectors have to make only one test.

Auto pollution monitors offer one of the greatest potentials for coated crystals. In addition to the instrumentation requirements of car dealers service stations, diagnostic centers and enforcement agencies almost 10 million cars are built each year. There are 80 million cars on the road. The possibility of using a built-in coated crystal auto pollution monitor for individual cars

for indicating high emissions offers tremendous potential to the crystal industry.

Although the automobile is the largest pollutor, the effect of pollution is on the individual. This raises a question whether we need protection to exposure. Exposure to radiation is monitored by individual dosimeters. A selective coated quartz crystal dosimeter has been developed for measuring exposure to hydrogen sulfide.

The challenge of this decade is "can we reduce our air pollution?" How we will monitor it will depend on the imagination of the scientist. The piezoelectric sensors have many desirable properties that can be utilized for pollution detectors.

We will require more research and development on the physics and mechanics of piezoelectric crystals as applied to pollution sensors.

We will need better manufacturing techniques for crystals to meet pollution requirements.

We will have to develop specific coatings for pollutants.

We will need low cost integrated circuit oscillators and temperature controllers.

Standards must be established for analyzer accuracy, precision response and selectivity.

This will results in a simple, efficient and low cost pollution analyzer that can be adapted to specific requirements.

THE STATE OF THE ART IN PIEZOELECTRIC SENSORS

W. H. King, Jr.
Esso Research and Engineering Company
Linden, New Jersey

Summary

Under carefully controlled conditions, AT cut crystals can detect 10^{-12} grams of deposit. Therefore, it is not surprising that quartz crystal resonators of varying descriptions are used as microbalances. This paper will focus attention to the many uses of crystals other than for filters and the control of frequency. The list of other uses includes the measurement of gas density, pressure, temperature, strain, oil vapor deposition, metal film thickness, dew point, interferometry, surface area, oxidation of elastomers, corrosion of metal, radiosonde humidity, polymerization, viscosity, vacuum, power, voltage, current, thermal conductivity, and volatility of asphalt. A number of sorption detectors have been developed for the measurement of particles, humidity, and hydrocarbon pollution in the atmosphere. Other sorption detectors now exist which are capable of detecting solvent vapors and other gaseous constituents.

Introduction

It is well known how troublesome contamination in a sealed quartz resonator can be when the object of the game is stability. Under carefully controlled conditions, AT cut crystals can detect 10^{-12} grams of deposit.[1] Therefore, it is not surprising that quartz crystal resonators of varying descriptions are used as microbalances. This paper will focus attention to the many uses of crystals other than for filters and the control of frequency. The list of other uses includes the measurement of gas density,[2] pressure,[3] temperature,[4] strain,[5] oil vapor deposition,[6] metal film thickness,[7] dew point,[8] interferometry,[9] surface area,[10] oxidation of elastomers,[11] corrosion of metal,[12] radiosonde humidity,[13] polymerization,[14] viscosity,[15] vacuum,[16] power, voltage, current, and thermal conductivity,[16,17,18] and volatility of asphalt.[19] A number of sorption detectors have been developed[16,20] for the measurement of particles, humidity, and hydrocarbon pollution in the atmosphere.[21] Other sorption detectors now exist which are capable of detecting solvent vapors and other gaseous constituents.[16]

Sorption Detectors

Sorption detectors are made by coating an electrode with a material that can sorb the gas in question. Thus, the more gas picked up, the heavier the electrode gets and the frequency decreases. Two crystals are used--one coated. The other is either sealed or uncoated and exposed to the same gas and temperature as the coated crystal. The audio frequency difference is used as the signal. Crystals without purposely applied coatings are slightly sensitive to moisture absorption in the manner shown by Figure 1. The small response shown is normally swamped out when a special coating is applied to the crystal. The type of coating material defines the performance of the detector as shown by Figure 2. Here the relative absorption for water on molecular sieves, polymers, etc., is shown for the low range (0-3% relative humidity). Molecular sieves are most sensitive in the 0-200 ppm range. The signal obtained from a detector with 20 KHz of molecular sieve coating will be 2,000 Hz at 100 ppm. On the other end of the scale are hygroscopic liquid coatings whose main virtue is good linearity but limited response at the low end.

A water detector is now commercially available using crystals coated with a hygroscopic polymer.[24] For air pollution, two instruments using sorption detector crystals are already developed. One is for ambient air hydrocarbon analysis where high sensitivity is important, and the second is for auto exhaust hydrocarbon analysis where speed and simplicity are important.

Most resonators for sensing use AT cut crystals from 4 to 10 MHz of various dimensions. Sorption detectors operate at ambient or at 50°C when constant temperature is required. Absolute frequency is not very critical since circuits usually contain two crystals which are trimmed electrically at some known or zero condition. One crystal is the sensor and the other a reference. The audio beat note between the two is the output signal. In a typical application a sorption detector sensor will receive 5 to 50 KHz of a coating over the center RF electrode. The larger the amount applied, the larger the sensitivity. The reference is nickel plated to match $\pm$ 1 KHz.

Sorption detectors can detect many kinds of vapors depending on the nature of the coating materials used. Table I shows the responses of typical units. Detectors No. 1 and 2 were designed to be used with gas chromatography. The response speed is the order of 40 milliseconds. They are also linear. By employing these two detectors and a thermal conductivity cell, one can instantaneously analyze the gas chromatograph effluent for boiling point, polarity, and concentration of eluent vapors.[16] Silica gel (No. 3) is an all-purpose detector because many materials are absorbed by the silica gel. Its response is a little slower than that of Detectors 1 and 2. Detector No. 4 is a hygroscopic polymer employed in a commercial water detector which is especially sensitive in the low range.[24] Detector No. 5 is not sensitive to water, but it is very sensitive to many kinds of solvent vapors. A leak detector was made employing this sensor. The detector is very useful in finding leaking fittings on apparatus where solvents or vapor are contained under pressure. Our detector is employed mainly to find leaks on gel permeation chromatographs and similar liquid chromatography equipment. The performance of a solvent leak detector is shown by Table II.

Integral Heater Sensors

By modifying one RF electrode into a resistance heater, a temperature sensitive AC or similar cut crystal becomes a transducer and can be used to measure voltage, current, power and thermal conductivity. In this application the heater is a wide, meandering, and closely spaced film having a resistance of 10 ohms. The other electrode is a normal RF electrode. Figure 3 depicts a typical integral heater crystal. The crystal can be AC cut to obtain a high temperature coefficient. Typical response to electrical current is shown in Tables III and IV. The sensor's parameters were F_r = 8924.97, F_{32} = 8932.75, R_1 = 13 ohms, Co = 14 PF, and heater resistance 12.8 ohms at 25°C, AC cut, 9/.5 in O.D. flat plate, wire clip.

The crystal microbalance has had several space applications such as measuring contamination buildup on the exterior of spacecraft.[22] A carbon dioxide, water and dust sensor system using crystals was recently developed for use on the 1975-76 Viking Mars lander. This work is not published.[23]

A sulfide dosimeter was developed using a silver electrode crystal in a protective but porous container. As sulfide in the air tarnished the silver, the frequency changed. If this and other possible personal dosimeters become accepted, a large market could result.

There is a potential for large-scale uses of crystals in the areas of air pollution measurements, radiosonde balloons, anesthesiology, solvent alarm systems, dosimeters and miscellaneous laboratory applications.

The radiosonde application could use 400,000 crystals per year. In this application one crystal could be used for temperature--one for humidity and a third for pressure. All three systems have been demonstrated to be superior to alternate ways of sensing these variables. What has not been done is developing the capability of making 100,000 complete disposable units including radio, multiplexer, antenna, sensors and container for about $35 each. The balloon and hydrogen are separate items.

References

1. Warner, A. W. and Stockbridge, C. D., "Vacuum Microbalance Techniques," Vol. 3, Plenum Press, New York, 1963, p. 55.
2. Pacy, D. J., Vacuum, 9, 261 (1960).
3. Hammond, D. L. and Benjaminson, A., IEEE Spectrum, V6, #4, 53-58 (April 1969).
4. Hammond, D. L., Adamo, C. A. and Schmidt, P., ISA Trans., 4, 349 (1965).
5. Mingins, C. R., Barcus, L. C. and Perry, R. W., Proceedings of the 16th Annual Frequency Control Symposium, pp. 46-76, April 25-27, 1962.
6. Doring, G. G., Marx, G. and Sauerbrey, G., Vakuum-Technik, Vol. 15, No. 10 (1966).
7. Oberg, P. and Logensjo, J., Rev. Sci. Instrm., 30, 1053 (1959).
8. Van Dyke, K. S., U.S. Patent 257,171 (July 4, 1945).
9. Blackburn, R. K., U.S. Patent 2,536,025 (January 2, 1951).
10. Wade, W. H. and Allen, R. C., J. Colloid and Interface Science, 27, 722 (1968).
11. Fischer, W. F. and King, W. H. Jr., Anal. Chem., 39, 1265 (1967).
12. Littler, R. L., U.S. Patent 3,253,219 (May 29, 1966).
13. Gjessing, D. T., Holm, C. and Lanes, T., Electronic Letters, 3, No. 4 (1967).
14. Kunze, D., Peters, O. and Sauerbrey, G., Ztschr. F. Angew. Physik, 22(2), 69075 (1967).
15. Rouse, P. E. Bailey, E. D. and Minkin, J. A., Proc. Am. Petrol. Indust. III, 30, 59 (1950).
16. Research/Development, April 1969, pp. 28-34 and May 1969, pp. 28-33.
17. King, W. H. Jr., U.S. Patent 3,478,573, November 18, 1969.
18. King, W. H. Jr., U.S. Patent 3,531,663, September 29, 1970.
19. King, W. H. Jr. and Corbett, L. W., Anal. Chem.
20. King, W. H. Jr., Anal. Chem., 36, 1735 (1964).
21. King, W. H. Jr., Environmental Science and Technology, V4, 1136-1141 (December 1970).
22. NASA Tech. Briefs 67-10205, 67-10012 (1967).
23. This work was done for Martin-Marietta Corp. by the Esso Research and Engineering Co., W. H. King, Jr. and G. M. Varga, Contract No. RCO-446067.
24. DuPont Instrument Products, Glasgow, Delaware.

Table I

Typical Data Obtained with 9 MHz Sensors

No.	Coating Material	Sensitivity, Hz/mol % Hexane	Benzene	Water	Temperature °C
1	DC-200 Silicone Oil	117	50	0.1	25
2	1,2,3-Tris-Cyanoethoxy Propane	2.3	43	200	25
3	Silica Gel	116	426	2,000	25
4	Hygroscopic Polymer*	0.0	0.0	300,000	50
5	Rubbery Polymer	75	138	0.3	39

* Sensitivity is shown for the 10 ppm range.

Table II

Performance of a Solvent Leak Detector
Consisting of a Polymer Coated 9 MHz Crystal

Solvent Vapor in Air	Sensitivity Hz/ppm (v/v at 25°C)	Meter Deflection for 1 µg/cc % of Full Scale
C_1	0.0001	0.002
C_2	0.0013	0.02
C_3	0.006	0.2
nC_4	0.019	0.7
nC_5	0.064	2.9
nC_6	0.30	16.
Benzene	0.50	25.
nC_7	1.15	72.
Toluene	1.95	112.
nC_8	4.25	303.
Ethyl Benzene	6.70	444.
o-Xylene	9.70	642.
nC_9	12.50	1,000.
Trichlorobenzene	75.0	8,600.
Perchloroethylene	5.0	518.
Chloroform	0.5	38.
Tetrahydrofuran	0.4	19.
Water	0.03	0.3

Ambient noise level is 0.1 Hz; response time is less than 1 second for hexane and 5 to 10 seconds for trichlorobenzene.

Table III

AC Cut Integral Heater Electrical Sensor

Heater Condition		Frequency	ΔF	Power
Volts	Amps	KHz	Hz	Watts
0	0	8924.973	0	0
0.6645	0.0505	8925.994	1,021	0.032
1.389	0.100	8929.465	4,492	0.139
2.097	0.1418	8934.833	9,860	0.297
2.725	0.1735	8941.181	16,208	0.472
3.155	0.1918	8946.103	21,130	0.604
3.607	0.2098	8955.039	30,066	0.842
4.395	0.2345	8962.695	37,722	1.030

Table IV

Response Speed of Electrical Sensor

Time Vs. Frequency after Applying 200 ma.

Seconds	Frequency	% Response	Sensor Temperature
0	8924.97	0	34°C
10	8940.06	66	111
20	8945.97	90	141
30	8947.37	97	149
40	8947.78	100	156
50	8947.98	100	157

FIGURE 1

CLEAN UNCOATED CRYSTALS
SHOW SMALL HUMIDITY RESPONSE

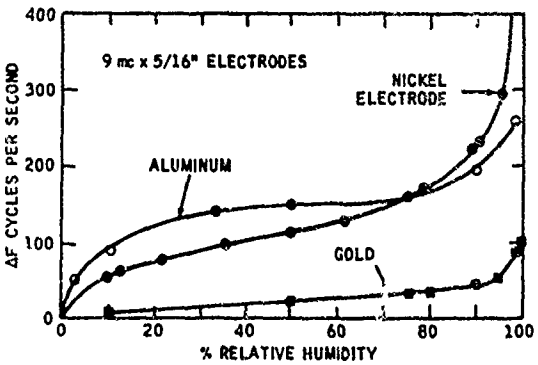


FIGURE 2

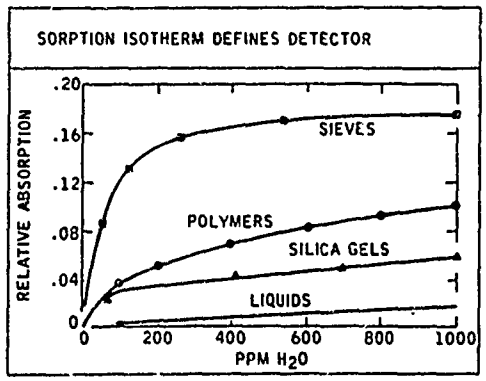
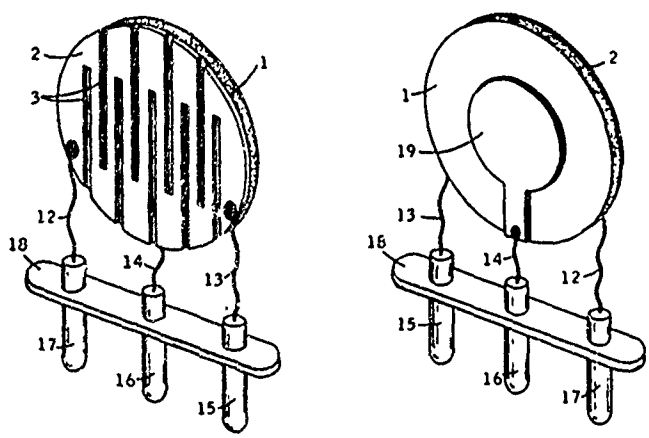


FIGURE 3



EVALUATION OF CULTURED QUARTZ FOR HIGH PRECISION RESONATORS

B. Capone, A. Kahan
Air Force Cambridge Research Laboratories (AFSC)
Bedford, Massachusetts 01730
and
B. Sawyer
Sawyer Research Products
Eastlake, Ohio 44094

Introduction

In an earlier paper,[1] the effects of electron irradiation on frequency and internal friction (Q^{-1}) were reported on oscillators fabricated from selected natural, fast rate Z-growth lithium nitrite and lithium carbonate doped, Z-growth electronic grade, swept Z-growth electronic grade, and, slow Z-growth, low impurity quartz. Measurements were performed on 5MHz, 5th overtone, high precision AT-cut resonators fabricated by Bliley Electric Co. The results showed that after exposure to 10 MeV electrons to a final dose of 1 Mrad(Si), swept Z-growth electronic grade quartz exhibited the highest frequency stability with no cessation of oscillation during irradiation. Also, little change was noted from the pre-irradiation internal friction spectrum. The small observed frequency change was positive and in the order of 0.02 ppm. The worst of the materials possessed frequency changes up to 10 ppm and gross changes in internal friction. In all cases there was evidence for a direct correlation between frequency deviation and changes in internal friction.

The possibility of fabricating resonators that exhibit no frequency change or stoppage during irradiation from an "as grown" material would alleviate the need for the expensive, poorly understood, and sometimes unreliable electrolytic (sweeping) process. This investigation concerns an evaluation of a cultured quartz combining several improved crystal growth features of the previously measured quartzes.

"Premium Q Grade" Quartz

This study describes the characteristics of "Premium Q grade" cultured quartz, from Sawyer Research Products. Rigid controls of pertinent growth parameters yielded material with Q values as high as 3×10^6. The distinguishing factors affecting the growth of this quartz start with the selection of good raw material. Lithium was included in the growth solution and a slow growth rate along with special controls on absolute temperature and temperature gradients between the dissolving and growth chambers was utilized. Thus, the smooth transition without overshoot from room temperature to operating temperature results in an improved heat-up sequence and an improved join between the feed and the new material.

Experiment

The experimental procedures for measuring internal friction and for the irradiation and frequency monitoring are as previously discussed.[1] Q^{-1} measurements were made using a transmission technique whereby the output amplitude of the resonator was displayed as a function of frequency.

Irradiations were performed at the AFCRL Linear Accelerator. Two resonator units were subjected to 10 MeV electrons in a 4.5 μsec. pulse to a dose of 1 Mrad(Si). The oscillator frequency and amplitude was monitored during each pulse. The limit of accuracy for the frequency measuring system is 0.02 ppm. After internal friction measurements were performed from 4.2 to 400°K, the dosage was increased to 10 Mrad(Si).

Frequency Results

The accumulated frequency offset of the irradiated resonators after 1 Mrad(Si) was 3Hz or 0.6 ppm and negative. A plot of accumulated offset as a function of dose is shown in Figure 1 along with the offsets for selected natural, lithium doped cultured quartz from Western Electric, and "High Q grade" cultured quartz also from Sawyer Research Products. Selected natural quartz at this dosage has a negative offset of about 20Hz. Offsets as high as 50Hz have been measured on unselected natural quartz in this laboratory whereas Poll and Ridgway[2] report values about 40Hz. The accumulated offset for Western Electric fast growth lithium doped cultured quartz is negative and about 40Hz, and High-Q quartz exhibits a positive offset of 18Hz. The best material measured to date is still swept electronic grade cultured quartz with a positive frequency offset of about 0.1Hz or 0.02 ppm.[1-3]

After each radiation pulse, Premium Q quartz resonators stopped oscillating for short time periods (10-26 seconds). Under identical experimental conditions of oscillator circuit, temperature control, and frequency monitoring techniques, neither swept electronic nor High-Q quartz oscillators stopped. However all natural and lithium doped quartzes did cease functioning for similar time periods (10-30 seconds). It is reasonable to assume that the inclusion of lithium in the growth solution may contribute to the stoppage of oscillations during irradiation. Consequently, Premium Q quartz is an improvement over all untreated quartz in terms of frequency stability, but not in terms of cessation of oscillations.

Internal Friction Results

Internal friction measurements on this material indicate Q values are in the range of 2.8 to 3×10^6 at normal operating temperatures as a 5MHz, 5th overtone, high precision AT-cut resonator. Under the same experimental conditions, typical values for similar resonators are in the range of 2.5×10^6. Investigations of internal friction from 4.2°K to 400°K plotted in Figure 2 as Q^{-1} vs temperature indicate that the region

of High-Q extends from 170°K to 400°K and is limited on the low temperature end of this range by coupling to other modes of oscillation. This coupling is inherent to AT-cut resonators and is associated with the angle of the cut.[1] Measurements were terminated at 400°K to prevent possible mechanical damage to the oscillator package (solder joints, supporting members, etc.).

At the low temperature end of the range the 50°K loss peak attributed to sodium impurities is small, there is a hump at approximately 30°K probably arising from phonon-phonon interaction, and the loss decline from 20°K to 4.2°K is another indication of High-Q ($\approx 2.5 \times 10^7$ at 4.2°K).

Figure 3 shows a comparison of the Q^{-1} spectra of the materials compared in Figure 1. Selected natural and fast growth lithium doped quartz exhibit the high temperature tail associated with alkali impurities in certain positions, this tail (actually the low end of the 750° loss peak) is absent in Premium Q quartz and High-Q quartz. The 50°K loss peak is small in selected natural and fast growth lithium doped quartz as well as in Premium Q quartz, while High-Q quartz, on the other hand, has a large 50°K loss peak. It is believed the lithium doping is responsible for this reduction. The growth rate in High-Q and Premium Q quartz was slow enough to prevent the formation of the high temperature tail but not slow enough to inhibit the growth of the 50°K loss peak in High-Q quartz. The phonon-phonon peak is not as pronounced in the other quartzes as in Premium Q. The low temperature decline found in both High-Q and Premium Q quartz is another signature of slow growth.

Figure 3 also shows a plot of post-irradiated Q^{-1} spectra. After irradiation, the internal friction of Premium Q quartz shows no significant changes. The very low temperature decline is unchanged, the small 50°K loss peak is further reduced and the overall background is just slightly above the pre-irradiation level. Natural and fast growth lithium doped quartz develop broad loss peaks in the 200° - 300°K temperature region and losses are also induced at approximately 10°K. The only appreciable change in High-Q quartz is the lowering of the 50°K loss peak.

Premium Q was developed to try to combine the best features of High-Q quartz and lithium doped quartz namely the absence of a 50°K loss peak, the elimination of the high temperature tail, and the prevention of the radiation induced band in the 200° - 300°K temperature region. To this end the product is successful and to a degree a better radiation hardened material than all but swept electronic grade quartz.

Characteristics Above 1 Mrad(Si)

Figure 4 shows an expanded plot of frequency offset of Premium Q with the dose to 5 Mrad(Si). The offset levels off between 0.7 and 0.9 x 10^6 Rads(Si) suggesting that saturation may be taking place. After maintaining a constant frequency output for 16 hrs. after the initial irradiation, the dosage was increased to 5 Mrad(Si). The frequency deviations now changed to a positive direction and reached a value only 0.7Hz below the pre-irradiation frequency. After each pulse the resonators stopped oscillating. Recovery times were up to 35 seconds for the resumption of oscillations. Full amplitude was reached after 5 minutes. After this irradiation Q^{-1} has been remeasured with the only noticeable change being a further reduction of the 50°K loss peak. Subsequent measurements of frequency offset to a dosage of 10 Mrad(Si) showed signs of partial annealing from the heating to 400°K during Q^{-1} measurements. The trend, however, was definitely toward positive frequency offsets after an initial level region.

Returning to Figure 2 we see that except for the almost complete elimination of the 50°K loss peak, changes in Q^{-1} after 10 Mrad(Si) are not significantly different from the changes after 1 Mrad(Si). The high temperature region of the Q^{-1} spectrum was again raised slightly and the small peak near 30°K is increased very little. It is interesting to note that up to 10 Mrad(Si) no saturation of the accumulated frequency offset has taken place.

Conclusions

Resonators fabricated from Premium Q quartz will function as high precision devices. The special precautions taken during its growth to control temperatures and starting materials has resulted in an improvement in the quality of cultured quartz. Litnium doping as found in Western Electric Cultured quartz reduces the 50°K loss peak and contributes to the nigh Q of the material. The slow growth process eliminates the higher temperature alkali losses and is responsible for the extremely high Q at lower temperatures. Although the inclusion of lithium leads to stoppage of oscillations during irradiation, the overall frequency stability is next best to swept electronic grade cultured quartz. It is the slight increase in background losses that is responsible for the small negative frequency offset during irradiations to 1 Mrad(Si).

The high Q value achieved over such a large temperature range indicates an approach to the practical if not the theoretical limit of Q for cultured quartz.

Acknowledgement

The authors express their appreciation to O. M. Clark for assistance in the experimental phase of the investigation.

References

1. B. R. Capone, A. Kahan, R. N. Brown, and J. R. Buckmelter, IEEE Trans. on Nuc. Sci., NS-17, 217 (1970).
2. R. A. Poll and S. L. Ridgway, IEEE Trans. on Nuc. Sci., NS-13, 130 (1966).
3. T. M. Flanagan and T. F. Wrobel, IEEE Trans. on Nuc. Sci., NS-16, 130 (1969).

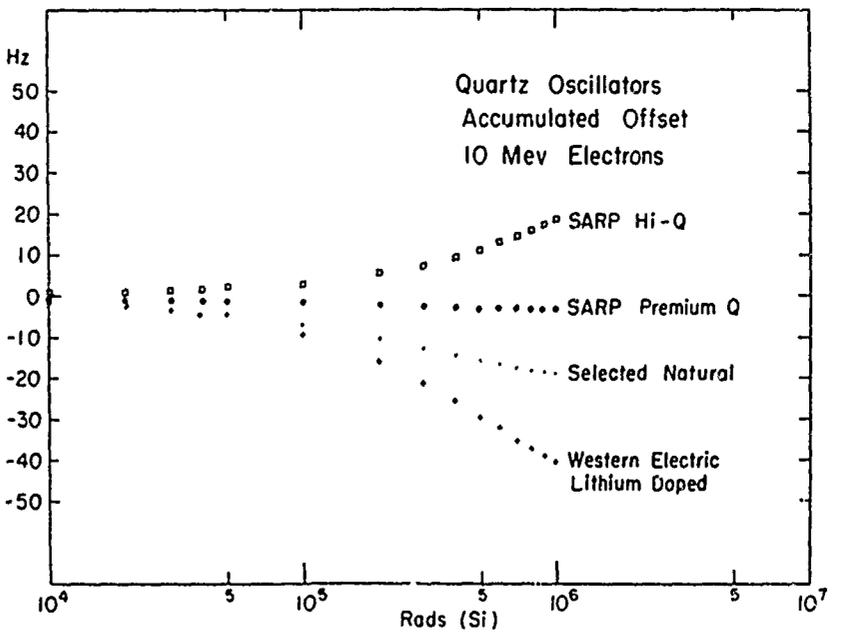


Fig. 1 - Accumulated Offset Oscillator Frequency as a Function of Electron Dose.

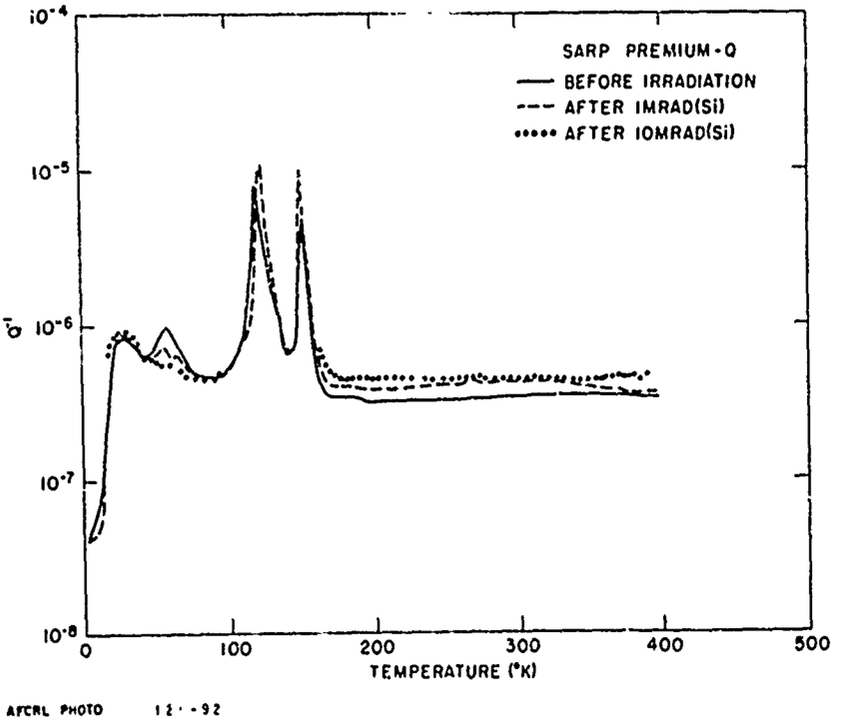


Fig. 2 - Internal Friction of Premium Q Grade Quartz vs. Temperature Before and After Irradiation.

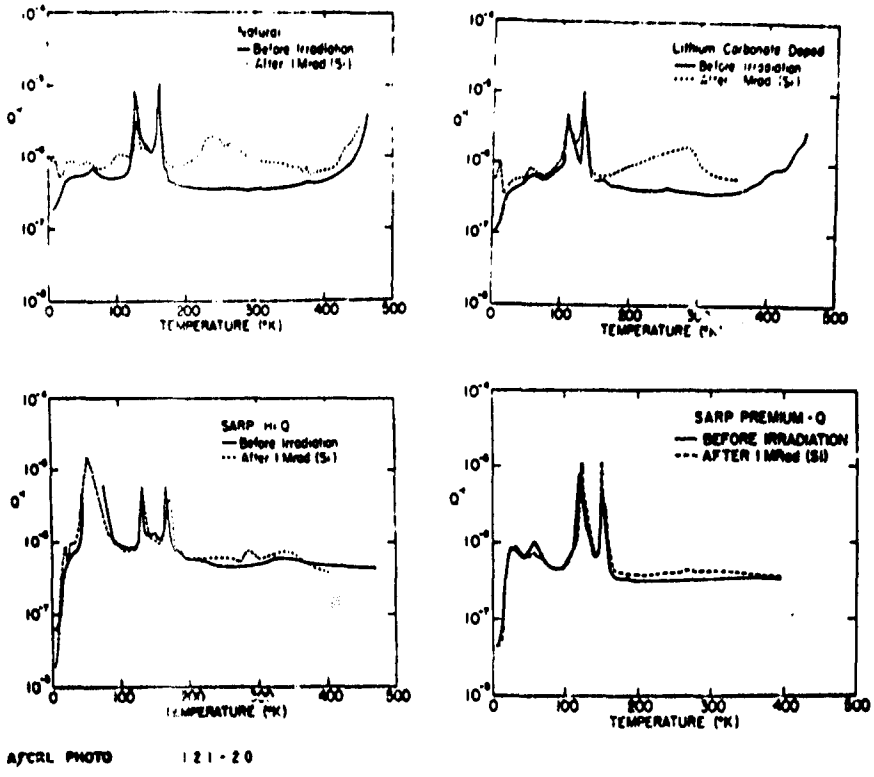


Fig. 3 - Internal Friction of Several Quartz Types vs. Temperature Before and After Irradiation.

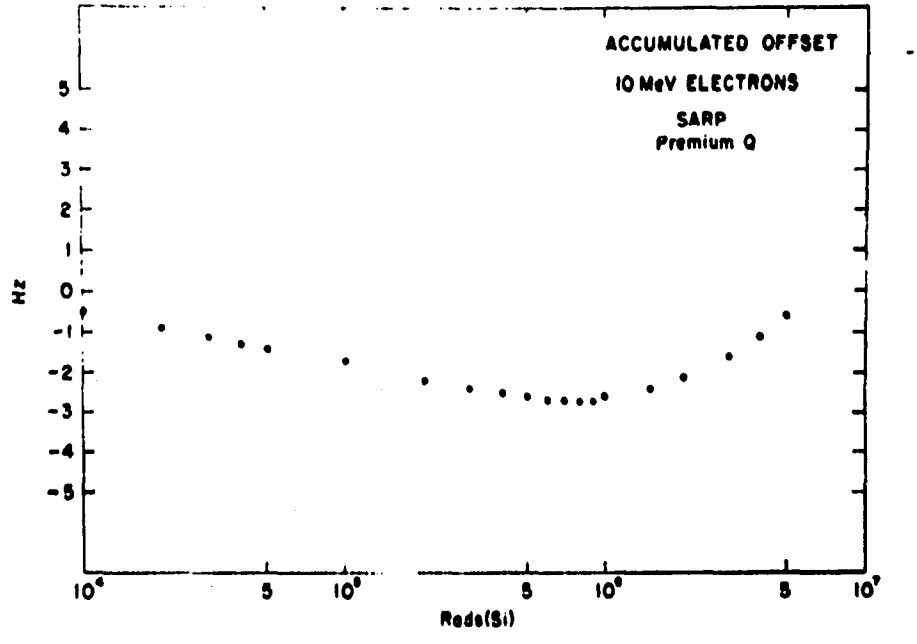


Fig. 4 - Expanded Plot of Accumulated Offset of Premium Q Grade Quartz to 5 Mrad(Si).

MEASUREMENT OF VIBRATION MODES OF PIEZOELECTRIC RESONATORS BY MEANS OF HOLOGRAPHY

Y. Tsuzuki, Y. Hirose, and K. Iijima
Yokohama National University
Yokohama, Japan 233

Summary

A holographic method of studying the vibration modes of piezoelectric resonators is described. This method was applied to the measurement of the vibration displacement distributions of the length extensional mode and the length-width flexural mode of rectangular quartz plates. Fairly good agreement was obtained between calculations and measurements, and it was verified that this method can be applied to the quantative measurement of the contour modes of vibration.

As an example of the application of this method to vibrating objects having a relatively complicated structure, measurement was made on the spurious modes of electromechanical filters made of several bending resonators coupled with each other. As a result, it became clear that there are two types of spurious modes which resemble the twist mode and the bending mode of a rectangular plate.

It is concluded from these results that the holographic method can be of great help in the vibration study of piezoelectric resonators.

Introduction

The holographic measurement method of vibration displacement distribution is one variation of the optical interferometric method. The observation of the vibration modes of quartz plates by means of the optical interferometric method was made by several workers about 40 years ago.[1] However, the adjustment of the measurement setup was very delicate in the previous method. Moreover, a resonator used for the experiment had to have an optical flat surface. After the invention of the gas laser, the observation of a vibrating metal disk having a diffuse surface was made by means of a holographic technique by Powell and Stetson,[2] and it was shown that the disadvantages of the classical interferometric method can be eliminated by the use of holography.

The authors applied the holographic method to various types of resonators, and verified that this method can be of great help in the vibration study of piezoelectric resonators.

Measurement Method

The measurement method used is illustrated in Figures 1 (a) and (b). First, an enlarged laser beam was projected onto a mirror and a resonator having a lapped surface. The reflected beam from the mirror acted as a reference beam, and a part of the scattered reflected beam from the resonator became a signal beam. The two beams reached a high-resolution photographic plate as shown in Figure 1 (a), and the interference pattern was recorded in the plate. The developed photographic plate is called a hologram. Next, the resonator was removed and the hologram was placed in the initial position as shown in Figure 1 (b). Then, the reference beam was diffracted by the hologram, and the original signal beam was reproduced. As a result, the three-dimensional reconstructed image of the resonator could be observed at the initial position through the hologram. This reconstructed image could also be photographed by a camera. The picture of the image was completely identical with the picture obtained directly by usual photography. Figure 2 (a) is an example of the reconstructed image of a resonator in a stationary state. This resonator is a 5° X-cut crystal unit for a lattice type crystal filter, therefore, the evaporated electrodes were divided at the center. The two dark lines in the lower part of the crystal surface in Figure 2 (a) are the shadows of the supporting wires. Another reconstructed image of the same crystal unit is shown in Figure 2 (b). In this case, the crystal plate was excited in its fundamental length extensional mode, and a pattern of dark fringes appeared in the reconstructed image. The relation between the pattern of dark fringes and the distribution of vibration displacement was discussed already by Powell and Stetson.[2]

When a resonator is vibrating, the signal beam receives phase-modulation by the periodic change of the optical path length. The amplitude of the phase change of the signal beam is given by

$$\phi(p) = \frac{2\pi}{\lambda} (\cos\theta_i + \cos\theta_r)\, U(p) \qquad (1)$$

where λ is the wavelength of the laser beam, θ_i or θ_r the angle between the direction of the vibration displacement and the incident beam or the reflected beam respectively, $U(p)$ the amplitude of the vibration displacement at an arbitrary point p on the plate surface. Then, the dark fringes appear on the plate surface of the reconstructed image where the following equation is satisfied.

$$J_o[\phi(p)] = 0 \qquad (2)$$

where J_o is the zero order Bessel function. The amplitude of the vibration displacement, $U(p)$, is zero at a node, therefore, the location of a node appears as the brightest portion in an image independently of θ_i or θ_r, the angle between the displacement direction and the incident beam or the reflected beam. If θ_i and θ_r are constant all over the surface of a plate, a pattern of dark fringes can be interpreted as a displacement distribution pattern, and U_n, the displacement amplitude at the location of the nth dark fringe counted from the brightest portion in the image, can be calculated from the following equation.

$$U_n = \frac{a_n}{(2\pi/\lambda)(\cos\theta_i + \cos\theta_r)} \qquad (3)$$

$$\approx \frac{(4n-1)\lambda}{8(\cos\theta_i+\cos\theta_r)} \qquad (4)$$

In Equation 3, a_n is the nth root of $J_o = 0$.

Experimental Apparatus

The experimental setup consisted mainly of a He-Ne gas laser, a collimator, an optical turntable, a mirror, and a camera with a standared lens or a telephoto lens. The collimator was used for the enlargement of the beam cross-section. The optical turntable was originally a part of an optical spectrometer, and was used for the adjustment of the direction of the incident beam upon the surface of the resonator so as to obtain a distribution pattern of the displacement component in a particular direction. A camera with a telephoto lens of 200 mm focal length was used for taking a picture of a reconstructed image from a relatively long distance so as to make θ_r, the angle between the displacement direction and the reflected beam, almost constant. As high-resolution photographic plates for making holograms, Eastman Kodak 649F plates were used.

Experimental Results

In order to evaluate the accuracy of the measurement of the vibration displacement in the surface direction of the resonator, the displacement distribution of a simple vibration mode was measured and compared with the calculated distribution. The length extensional mode of a thin -18.5° X-cut quartz plate was used as this mode is very close to an ideal one-dimensional vibration. The dimensions of the quartz plate were 40.0 by 3.75 by 0.80 mm, and the resonance frequency of the extensional mode was 63.9 kHz. Figures 3 (a) and (b) are the reconstructed images showing the displacement distributions with the crystal drive currents of 1 mA and 2 mA. The positions of the dark fringes in these images were measured and plotted in Figure 3 (c). The calculated distributions are also shown in the same figure as the solid lines. A satisfactly agreement was obtained, and the estimated error was less than 5%.[3]

Next, the capability of the measurement of the displacement component in a particular direction was examined by measuring the distribution patterns of the lengthwise displacement and the widthwise displacement of the second overtone length-width flexural mode of a -18.5° X-cut plate. The dimensions of the plate were 40.0 by 3.75 by 0.80 mm, and the resonance frequency of the second overtone length-width flexural mode was 32.0 kHz. The quartz plate was supported at the center, and the evaporated electrodes were divided skew-symmetrically for excitation at its second overtone mode. The reconstructed images of this mode are shown in Figures 4 (a) and (b). For the measurement of (a) and (b), the incident beam was projected upon the plate surface from the length direction or the width direction respectively. In Figure 4, the calculated patterns are shown for comparison, and in this case also the agreement was fairly good.[3]

As an example of the application of the holographic method to vibrating objects having a relatively complicated structure, the measurement was made of various modes of electromechanical filters made of several bending metal resonators coupled with each other. Figure 5 (a) is the reconstructed image of a 4-element mechanical filter in a stationary state. The dimensions of the bending resonators were 70 by 4.0 by 1.7 mm. The center frequency of the pass band of the filter was 1860 Hz, and the band width was 100 Hz. On the other side of the two resonators at the two ends of the filter, a ferroelectric ceramic wafer was attached as the electromechanical transducer. Figure 5 (b) is the reconstructed image of the same filter showing the displacement distribution in the pass band. In the case of the mechanical filters, the incident beam was projected from the perpendicular to the filter surface so as to measure the vibration displacement perpendicular to the surface. From Figure 5 (b), it can be easily understood that the resonators were coupled with each other at the nodes of the bending mode. Figures 6 (a) and (b) are the reconstructed images of the spurious modes which appeared in the lower side and the upper side of the pass band. From the locations of the brightest portions and the distribution of the dark fringes, it became clear that these spurious modes resemble the fundamental mode and the second overtone mode of the twist vibration of a rectangular plate as shown in Figure 6. It was found also that there is another series of spurious modes which resembles the bending vibration of a rectangular plate. In this series, the deformation of the filter occurs mainly by bending deformation of the coupling metal rods. From these results, a way of rejecting spurious responses was established.[4]

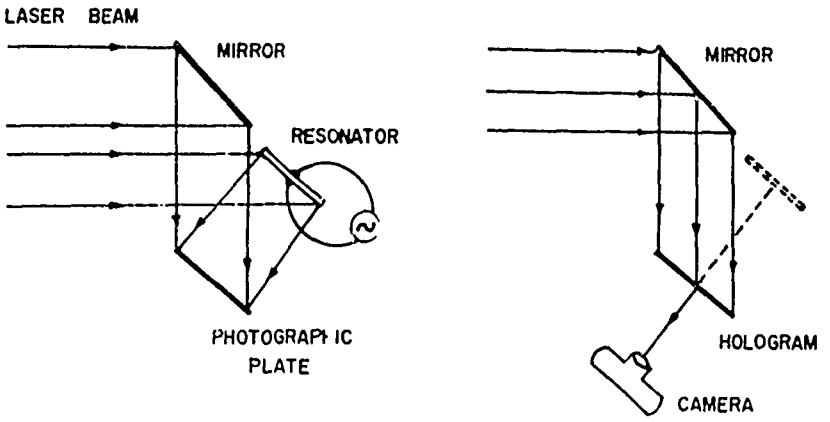
Conclusion

It is concluded from the experimental results that the holographic measurement method of vibration displacement distribution has the following advantages in comparison with the classical optical interferometric method.

(1) The quantitative measurement of the displacement component distribution resolved in a particular direction can be made on the so-called contour modes of vibration such as the extensional mode and the flexural mode.
(2) The measurement of the displacement distribution can be made also on the vibration modes of the composite vibrating systems of planar structures such as the electromechanical filters made of several bending resonators coupled with each other.
(3) It is not necessary to polish the surface of a resonator to optical flatness. The measurement can be made on a resonator with the usual lapped surface.

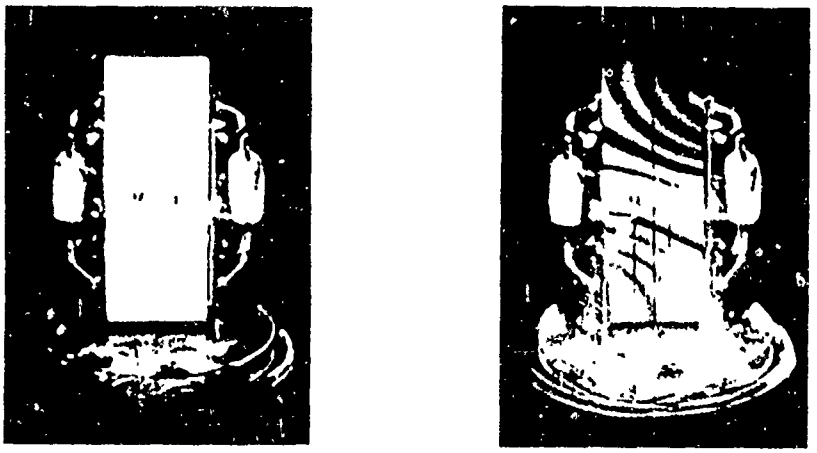
References

1. H Osterberg, "An interferometric method of studying the vibrations of an oscillating quartz plate", J.Opt.Soc.Amer., 22, p.19 (Jan. 1932)
2. R.L.Powell and K.A.Stetson,"Interferometric vibration analysis by wavefront reconstruction", J.Opt.Soc.Amer., 55, p.1539 (Dec. 1965)
3. Y.Hirose, Y.Tsuzuki and K.Iijima, "Measurement of contour vibrations of quartz plate by the holographic technique", Trans.IECE of Japan, 53-A, p.322 (June 1970)
4. Y.Tsuzuki, Y.Hirose, S.Takada and K.Iijima, "The holographic investigation of the spurious modes of the mechanical filters", Trans. IECE of Japan, 54-A, p.155 (March 1971)



(A) HOLOGRAM FORMATION (B) RECONSTRUCTION

FIG. 1 HOLOGRAPHIC MEASUREMENT METHOD OF VIBRATION MODE



(A) STATIONARY STATE, (B) EXTENSIONAL MODE

FIG.2 RECONSTRUCTED IMAGES OF A 5°X-CUT CRYSTAL UNIT

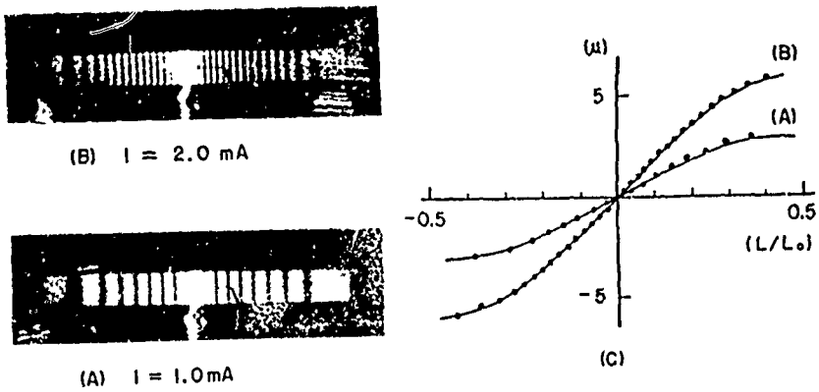


FIG.3 DISPLACEMENT DISTRIBUTION OF THE EXTENSIONAL MODE OF A -18,5° X-CUT PLATE

(A), (B) RECONSTRUCTED IMAGES

(C) MEASURED AND CALCULATED VALUES

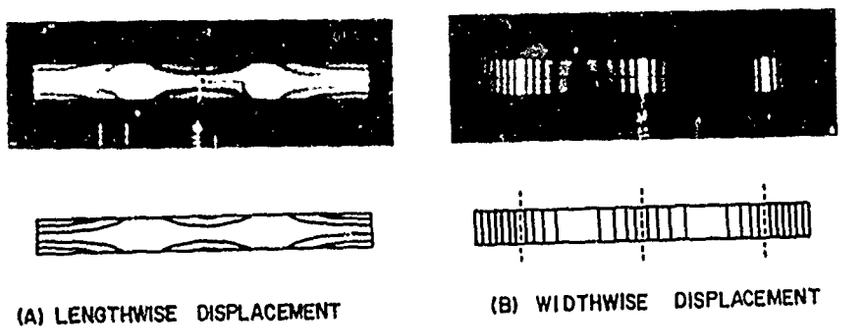
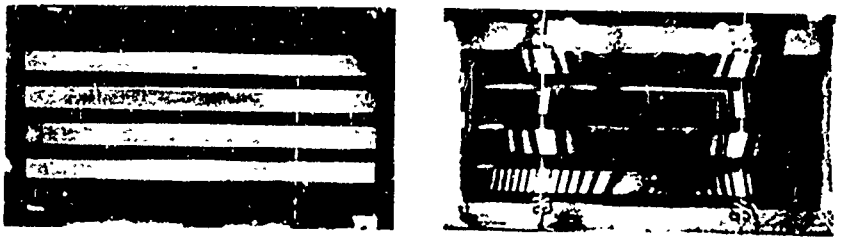


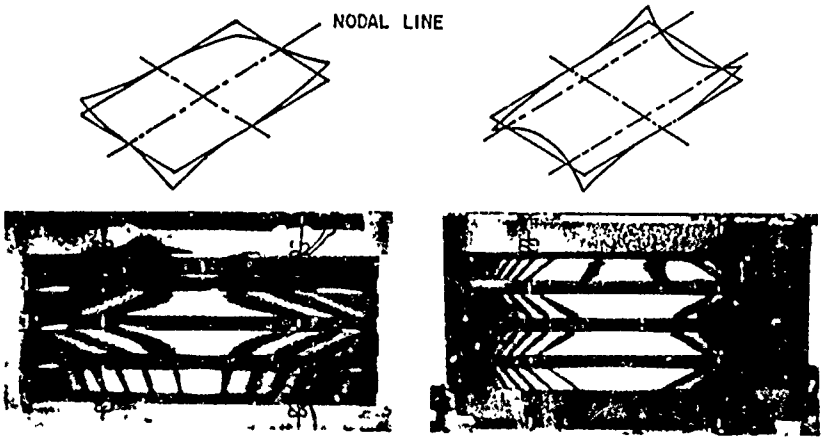
FIG.4 RECONSTRUCTED IMAGES AND CALCULATED PATTERNS OF THE SECOND OVERTONE LENGTH-WIDTH FLEXURAL MODE



(A) STATIONARY STATE

(B) VIBRATION MODE IN THE PASS BAND (1860 Hz)

FIG. 5 RECONSTRUCTED IMAGES OF A 4-ELEMENT MECHANICAL FILTER



(A) FUNDAMENTAL TWIST MODE (1439 Hz)

(B) SECOND OVERTONE TWIST MODE (2552 Hz)

FIG. 6 SPURIOUS MODES OF A MECHANICAL FILTER

STUDY OF FREQUENCY CONTROL DEVICES IN THE SCANNING ELECTRON MICROSCOPE

R. J. Gerdes and C. E. Wagner

Georgia Institute of Technology
Atlanta, Georgia 30332

Summary

Resonating quartz crystals which were actual frequency control devices were studied by a novel technique, scanning electron microscopy. During resonance various modes of thickness-shear, face-shear, flexure and other coupled modes became directly observable in the scanning electron microscope (SEM). The influence of surface defects on the resonance behavior of the quartz crystals was studied and the lattice strain associated with resonance could be assessed. Approximate values for the amplitudes of face-shear and thickness-shear could be obtained directly from the scanning electron micrographs.

The resonators consisted of SL- and AT- cut quartz slices commonly used in frequency control devices. The quartz surfaces did not have any conductive coating other than conventional thin film electrodes. The experiments were carried out in the vacuum of the SEM, usually 10^{-6} to 10^{-4} torr.

When the crystals were resonated at frequencies ranging from 265 KHz to 15 MHz characteristic patterns of stark black and white contrast were obtained. Bright areas on the scanning electron micrographs indicated the region of an antinode of displacement, whereas areas which appeared dark or black in the scanning electron micrographs coincided with the nodes of displacement.

The degree of black and white contrast was affected by defects in the crystal surface and by the drive level of the oscillator. Y- modulation of the secondary electron signal provided a suitable method of measuring contrast as a function of drive level and of defects.

While the observed contrast was indicative of the various states of electrical polarization on the crystal surfaces, details of surface topography also gave a direct indication of the amplitudes of displacement in the case of face- and thickness-shear. The direction and range of motion of various topographical features could actually be observed in the SEM as a function of drive level and frequency of excitation.

Key Words

Scanning Electron Microscopy, Quartz Crystal Resonators, Voltage Contrast, Vacuum.

Introduction

The scanning electron microscope (SEM) provides a novel experimental approach to the study of piezoelectric materials which are actual frequency control devices. A finely focused electron beam is scanned over the specimen surface in a square raster and produces x-rays, heat, cathodoluminescence, backscattered primary electrons and, in particular, secondary electrons. The usually recorded scanning electron microscope images are produced primarily by backscattered and secondary electrons. The effective secondary electron yield is a function of a number of factors, such as energy of the incident electron beam, specimen surface topography, and specimen surface potential. Insulators, such as quartz, have secondary electron yields which may be as high as 20 secondaries per incident 20 KV primary electron. Rapid depletion of electrons in the neighborhood of the specimen surface produces "charging", often positive charging relative to the incident electron beam. The secondary electron yield for metals by comparison is much lower, frequently it is less than 1. More favorable conditions are obtained for the study of insulators when low (1 to 2 KV) incident beam voltages are used. Under such conditions the secondary yield approaches approximately 1 and "charging" is generally not observed. It is of course understood that specimen surfaces discussed in this context do not have the conductive surface film which is usually applied for SEM studies of non-conducting materials.

Surface topography contributes significantly to the contrast (i.e., secondary electron yield as a function of surface detail) usually observed on scanning electron micrographs. But on quartz surfaces which do not show much surface detail at magnifications below 5,000 x, this contrast mechanism is of little importance, especially since the greater part of the present studies was conducted at magnifications of only 20 or 10 x. The contrast observed on resonating quartz crystals is therefore not produced by specimen surface topography but by variations in surface charges which are formed during resonance. Electrically charged surfaces significantly affect secondary emission characteristics, as was demonstrated by Everhart[1] and Nixon[2]. This voltage contrast has been particularly useful for investigations of integrated circuits and semiconductors. The direct observation of electric fields with the SEM was also reported by Spivak and coworkers[3] who investigated quartz and lithium niobate resonators which had been coated with germanium. Gerdes and Wagner[4] demonstrated that it is possible to study quartz crystal resonators which are actual frequency control devices without any conductive surface coating. As discussed above it is, however, advisable to utilize an incident beam voltage of only 1 or 2 KV to avoid "charging".

Ultrasonic vibrational modes in quartz crystal resonators have also been studied by a variety of other methods. The probe method[5] has led to information about surface polarization of selected quartz crystals. Light optical methods[6] have contributed information on surface displacements of resonating crystals. Images of strain variations along their various crystallographic directions have been produced by x-ray diffraction topography.[7]

In the present work it is shown that quartz resonators which are actual frequency control devices may be studied in the conventional SEM without the usually required specimen surface preparations. Compared to several of the above mentioned methods the studies are carried out in the vacuum of the SEM, limited only by the type of vacuum pumps used in the microscope. The image of the

resonating quartz crystal is obtained instantaneously in the SEM. Furthermore, quantitative analysis of the observed contrast and therefore also of strain variations at the specimen surface is possible by suitable signal processing. The possibilities of signal processing represent probably the real advantage in the use of the SEM.

At high magnifications it is possible to display directly surface displacements parallel to the surface as a function of drive level various surface defects, and various modes of resonances.

Experimental Procedures

Conventional SL- and AT-cut quartz crystal resonators were used in these studies. The SL-cut crystal was 10 mm x 4 mm x 0.7 mm in size. In the case of the AT-cut resonators the crystals consisted of circular plano-convex quartz slices of 1/2" diameter with an 8" radius of curvature for the convex side. Annular or circular aluminum electrodes were attached to the centers of the AT-cut crystals. In one case a thicker crystal with a z-shaped electrode was utilized. The SL-cut resonator was completely covered with Ag-Al thin film electrodes, except for the crystal edges. The resonators were driven by a Rohde and Schwartz crystal synthesizer or a crystal impedance meter at levels ranging from 2 to approximately 10 milliwatts. The experiments were carried out in a Stereoscan SEM at 1 or 2 KV. In order to obtain magnifications as low as 10 x the specimen tilt mechanism had been removed from the specimen stage. The emissive mode of operation was used throughout this study. Magnifications of 10 x were used, unless otherwise indicated by a micron marker.

Results

(a) SL-Cut Resonators

This type of resonators had previously been studied by methods of x-ray diffraction topography.[8] During resonance at 455 KHz the face-shear motion of this resonator was related to the motions of three principal blocks as shown in Figure 1. The a and b regions had

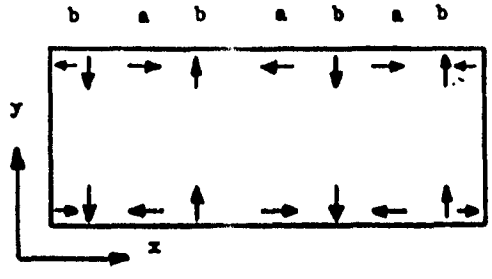


Figure 1. Location of nodes and antinodes of displacement in SL-cut crystal as determined from x-ray diffraction topography.

been interpreted as antinodes of displacement in the x- and y-directions, respectively, with nodes of displacement occurring along the crystal edges at position b. The dormant resonator is shown in Figure 2a. The edge of the crystal slab had not been coated with the electrode material. Because of the thin-film coating, voltage contrast during resonance was expected only along the edges. Figure 2b shows the top position of the crystal during resonance at 455 KHz, Figure 2c the center portion. The dark areas occurred approximately at the position of b along the crystal edge of Figure 1 and, therefore indicate the position of a node of displacement in agreement with the x-ray results. The same conclusion can also be reached from considerations of the secondary emission behavior of the quartz surface during ultrasonic excitation. During resonance the electric polarization is zero at an antinode of displacement. Secondary electron emission from an antinode should therefore be about the same as from a crystal surface which is not being resonated. The bright areas are therefore the locations of the antinodes of displacement. The electric polarization is at a maximum during resonance at a node of displacement where during the first few cycles of resonance, when the polarization is at its negative maximum, more secondary electrons are emitted than at an antinode. The result should be a net deficiency of electrons relative to the antinodes, positive charging and a

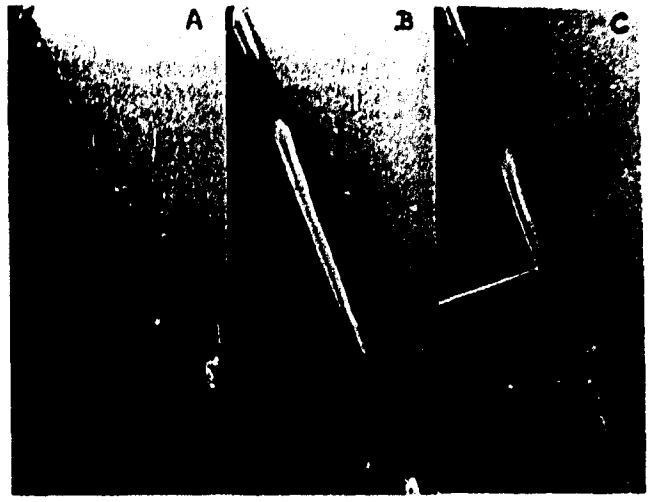


Figure 2. SL-cut resonator; (a) dormant, (b) during resonance at 455 KHz, top portion; (c) center portion.

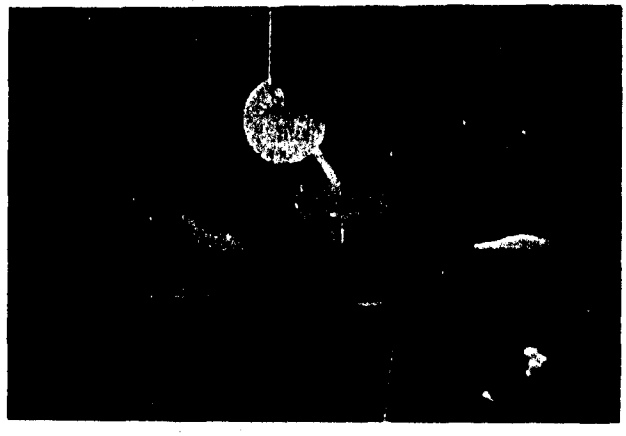


Figure 3. SL-cut resonator with portion of electrode etched away.

lower secondary electron yield. The nodes of displacement will therefore appear dark or black on the scanning electron micrographs. This relationship between positive charging and low secondary electron yield is consistent with observations by other authors.[9]

In Figure 3 a portion of the electrode material was etched away and a rather complex pattern appeared during resonance. While it was expected that the partially etched electrode would add to the complexity of the resonance pattern, the occurrence of a resonance pattern which was different along the crystal edges (Fig. 2) was surprising.

(b) AT-Cut Resonators

A portion of a dormant AT-cut resonator is shown in Figure 4. Figure 5 shows a 5 MHz crystal with annular electrode resonated at 5.221 MHz, the lowest frequency response that could be obtained with this crystal. The center portion shows a node of resonance which is probably due to thickness shear, whereas flexure modes probably produced the pattern outside of the annular electrode. The white area (arrow) outlines the shape of the bottom electrode. This one and other micrographs show that certain features, in particular the electrode material, visible through 1/2 mm and more of quartz. A unique explanation for this anomalous transmission is not available at present.

Figure 6 is the y-modulated micrograph of Figure 5. Bright areas in Figure 5 appeared as peaks in Figure 6. These peaks provide a direct means for quantitative contrast determinations. Assuming that there is a direct relationship between electric polarization and lattice strain during resonance then relative strain variations of 2 to 3 orders of magnitude can be determined across the surface of the resonators.

Figure 7 shows a resonance pattern (5.271 MHz) which was apparently caused by flexure modes only. It is interesting that these flexure modes are visible inside and outside of the annular electrode.

At 5.423 MHz (Figure 8) a pattern due to thickness shear is visible inside the electrode ring while flexure modes occurred outside of the ring. This pattern, as the patterns observed on all the other micrographs appeared on the SEM screen with a precision of at least a few Hz. Frequently the precision of pattern appearance was used to tune the crystal synthesizer to one of the resonance frequencies.

In Figure 9 the crystal was resonated at 5.428 MHz and the thickness shear pattern rotated through 60°. Another rotation through 45° occurred at 5.530 MHz (Figure 10). Resonance at 5.606 MHz produces a further 45° rotation of the thickness shear pattern (Figure 11).

In Figure 12 a surface defect on the center portion of a MHz AT-cut crystal with circular electrode is seen at rest. In Figure 12a,b,c and d the crystal was resonated at 5 MHz and increasing drive levels. Currents of 20, 30, and 40 μAmps were measured. During resonance surface features move in the direction indicated by the arrow and become elongated. As the distortion (elongation) of a given surface feature is twice the amplitude of the thickness-shear motion, the amplitude of motion can be determined. In the present case the amplitude of motion was found to be approximately 2000Å ± 500Å at 40μA. The precision of this measurement could be significantly improved by the use of time-resolved microscopy.[10] Studies in this area are in progress.

The micrograph in Figure 13 shows a portion of a crystal with a z-shaped electrode. The crystal was obtained from the laboratory of the late W. Ianouchevsky. The device was a 5 MHz crystal and the pattern in Figure 13 is that of the fundamental mode of resonance at 4.995 MHz.

The pattern in Figure 14 was obtained at 5.205 MHz. Two black bands appeared on both sides of the electrode related to each other by a 180° rotational symmetry operation.

Figure 15 is a micrograph of the same resonator as in Figure 14, but at resonance of 5.349 MHz. Five black bands have appeared on either side of the electrode. They seemed to be related to each other by some form of minor operation. But there were numerous features which seemed unique at a given area. The black bands have a white line somewhere close to the center of the band. Black lines also appear in the bright areas. Small surface irregularities which were randomly distributed over the entire surface of the resonator, appeared to have little effect on the positions of the bright and black bands. Thus, the image detail in these bright and black regions seems to be directly related to the respective mode of resonance which was used to excite the crystal.

Figure 16 was obtained through y-modulation of Figure 15. Contrast variations in the bright areas, in particular, stand out clearly in the micrograph.

In Figure 17 the drive level was increased approximately by five times compared to the drive level of Figure 15. As was to be expected, an even lower secondary electron yield was to be expected in the neighborhood of black bands, the nodes of displacement. The strong positive charging in these areas overshadows a l detail that had previously been visible. The expansion of the positively charged areas is especially noticeable in the y-modulated micrograph of Figure 18. The white line in the black bands appeared visible as an extended range of hills.

Resonance at 5.654 MHz produced the band structure in Figure 19. A total of ten bands was visible on either side of the electrode. The tenth band was just barely recorded on the lower right hand corner of the micrograph. The y-modulation micrograph of Figure 20 shows a number of interesting features. First of all the white line inside the black bands are quite noticeable as they are also in Figure 19. Then the contrast variations along the surface of the neutral electrode are of great interest. The top electrode of this crystal (and of all the others shown in this study) was grounded. The contrast observed in the SEM should therefore be uniform. Since this is not the case other factors, such as positioning of the crystal specimen relative to the secondary collector grid, were found to be critical in these experiments.

Discussion

The appearance of the standing wave pattern on the cathode ray tube of the SEM during ultrasonic excitation of the crystals in their various modes of resonance was instantaneous and sensitive to within a few Hz, probably surpassing the sensitivity of the Rohde and Schwartz crystal synthesizer. Excitation close to or in the neighborhood of a resonance frequency resulted either in no resonance pattern at all or, if the crystal had just gone through resonance, in a pattern with coviously diminished contrast and with decay-like features.

The various resonance patterns which were obtained changed characteristically from one resonance frequency to another. There was never any doubt about recognizing a particular pattern; therefore, it is concluded that there is a direct correlation of the SEM patterns which were produced by specimen surface polarization with actual bulk-effect resonance. The positions of the nodes and antinodes of polarization on the surface of the resonators, however, may have been affected by surface defects or the geometry of the specimen - secondary collector arrangement.

Secondary electron emission in the bright areas of resonating crystals was found to be approximately the same as on the surface of dormant crystals. The same or similar emission in these areas would indicate an area of zero surface polarization and therefore an antinode of displacement. It follows that the black areas are the nodes of displacement. The experimental proof for this interpretation will be sought in the utilization of time-resolved microscopy.

Charging, generally observed on the surface of insulators, such as quartz crystals, was no problem because

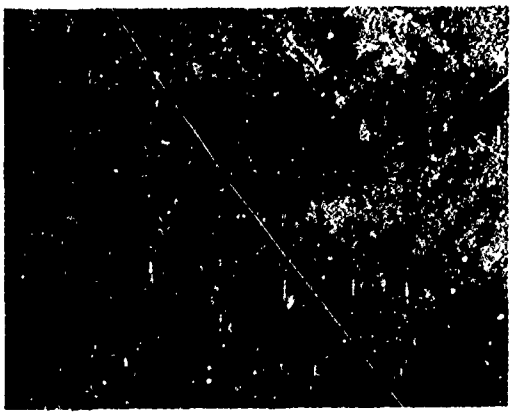


Figure 4. Dormant AT-cut resonator. Top half of micrograph shows portion of electrode, bottom half portion of quartz surface.

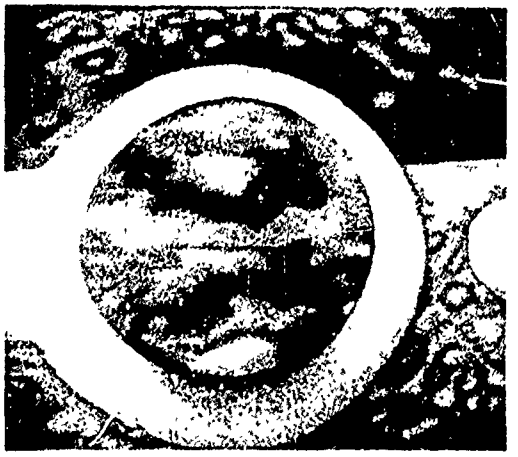


Figure 7. AT-cut resonator at 5.271 MHz.

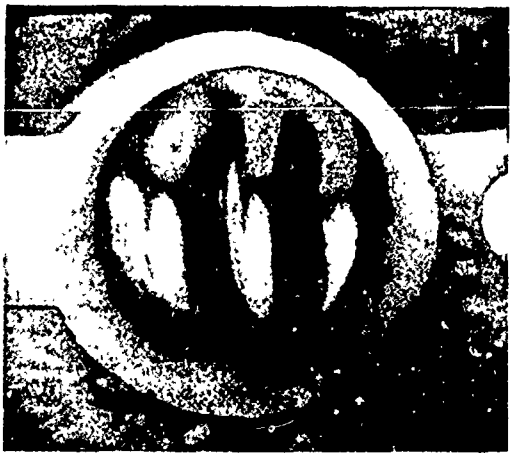


Figure 5. AT-cut resonator with annular electrode resonating at 5.221 MHz.

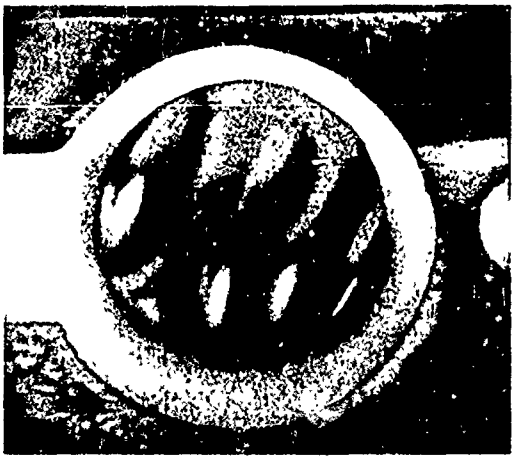


Figure 8. AT-cut resonator at 5.423 MHz.

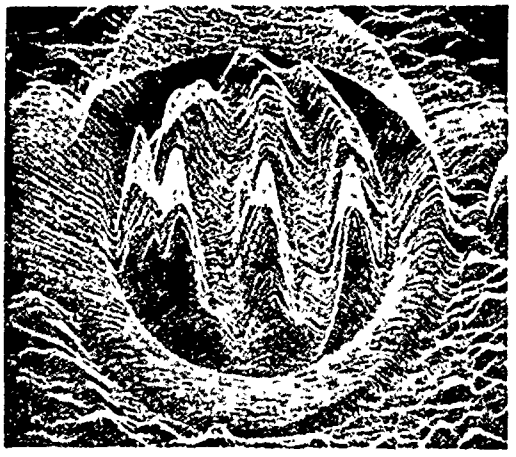


Figure 6. Y-modulated micrograph of Figure 5.

Figure 9. AT-cut quartz crystal during resonance at 5.418 MHz.

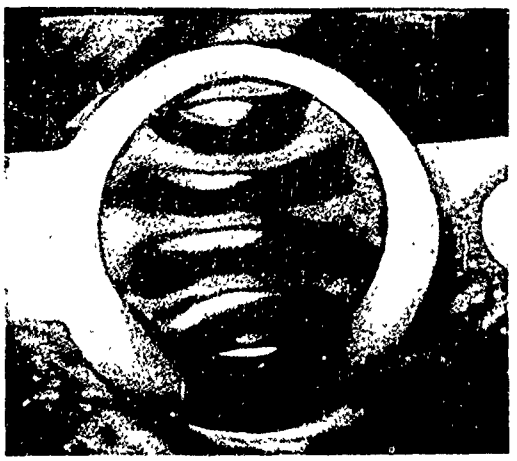


Figure 10. AT-cut crystal during resonance at 5.530 MHz.

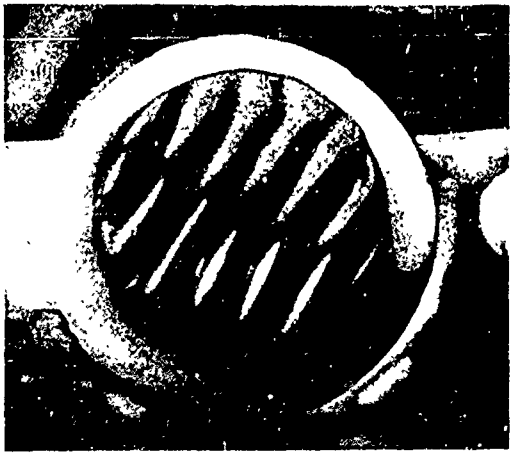


Figure 11. AT-cut crystal during resonance at 5.606 MHz.

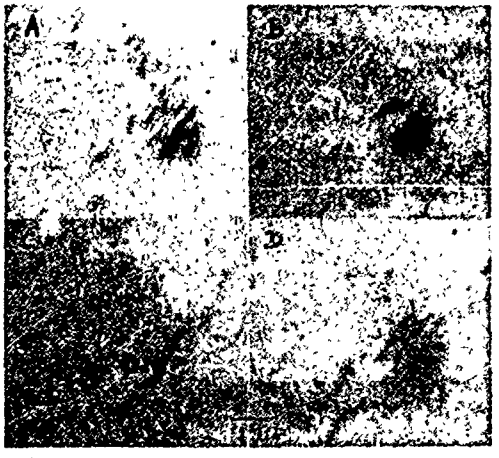


Figure 12. Surface defect on center portion of circular electrode of AT-cut crystal; (a) dormant crystal, (b) drive level of 20 30 (c), and 40 μAmps (d).

Figure 13. AT-cut device with z-shaped electrode during resonance at 4.995 Mhz. The diagonal band is a portion of the z-shaped electrode.

Figure 14. AT-cut device resonating at 5.205 MHz.

Figure 15. AT-cut crystal resonating at 5.349 Mhz.

Figure 16. Y-modulation micrograph of resonance mode in Figure 15.

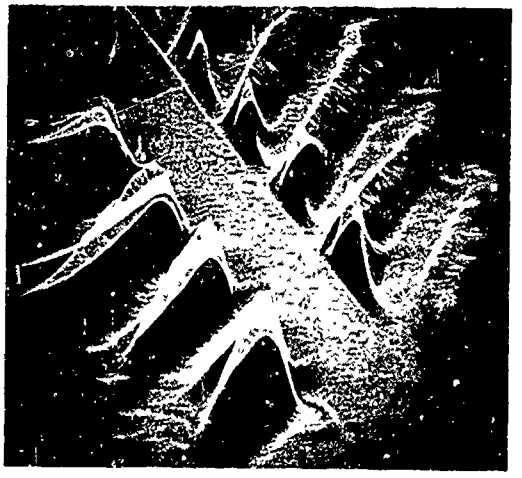


Figure 18. Y-modulation micrograph of resonance mode in Figure 17.

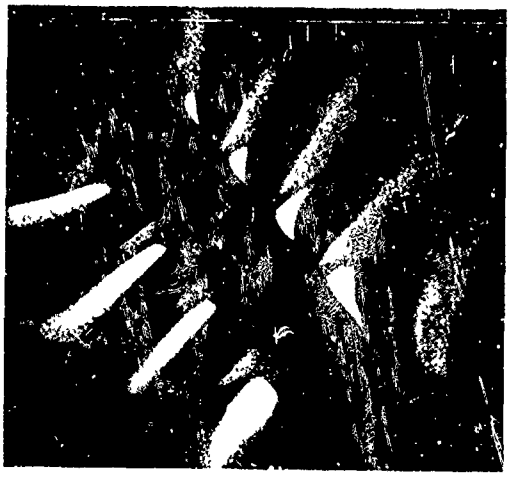
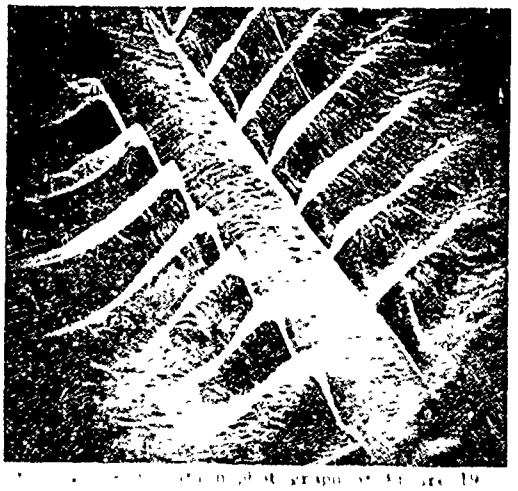


Figure 17. Resonance frequency is [illegible] Figure 16 [illegible] drive level [illegible] is [illegible] peak-to-peak [illegible].

Figure 19. AT-cut device resonated at 5.654 MHz.



[illegible] micrograph of Figure 19.

incident beam voltages of 1 or 2 KV were used in these studies. Contamination, by comparison, was a slight problem during long periods of observation. In electron microscopes contaminating layers are produced through interaction of the electron beam with the residual gases on the specimen surface. When this layer is thick enough the conductance of the quartz surface becomes too high for resonance pattern to be observable. The quartz resonators were therefore frequently cleaned in chromic acid prior to the SEM studies.

When the SEM was used in the usual mode of operation (frame speed of raster approximately 1 sec., line speed 0.004 sec.) the patterns were at once visible. Recording of the pattern with the commonly used 100 sec. frame speed and 0.1 sec. line speed, initially resulted in micrographs with very poor contrast. Recording times of 10 sec. frame and 0.01 sec. line speed resulted in acceptable micrographs. Extended sojourn times of the electron beam per unit area apparently tended to equalize variations in surface polarization. For future work it will therefore be necessary to determine quantitatively the incident electron beam current per unit area and time in such a way that optimum results for observation and recording are obtained.

Conclusions

The scanning electron microscopes provides an elegant and fast technique for the study of resonating frequency control devices. The appearance of the resonance pattern is sensitive to within a few Hz but may even exceed the sensitivity of the crystal synthesizer which was used in this study.

The studies are always carried out in the vacuum. The only limitations, if any, are the pumping systems which are commonly used in SEM's (10^{-6} torr). It is therefore, in principle, possible to carry out studies as a function of gas adsorption or temperature while the resonance behavior of the quartz crystals is under observation.

The amplitudes of face-shear and thickness-shear in SL- and AT-cut crystals can be measured directly from the motion of surface structures. These measurements will be significantly improved when stroboscopic methods methods are utilized.

The various modes of motion on the crystal surface during resonance are characteristic and easily distinguished on scanning electron micrographs. Modes of flexure, faces- and thickness-shear are directly observable.

The possibilities of signal processing in the SEM are especially attractive. Y-modulation provides a direct method of determining quantitatively signal intensities of various surface regions, and therefore various degrees of surface polarization, relative to each other.

The scanning electron microscope is unique because it combines a multitude of experimental capabilities which previously had needed a variety of different instruments. In the present study only a few of the SEM's capabilities have been utilized, other equally important ones have yet to be explored. These other possibilities include Auger analysis, energy and wave length dispersive x-ray analysis, and cathodoluminescence studies. Auger analysis would be especially useful for studies of the resonance behavior of quartz crystals as a function of gas adsorption. X-ray fluorescence analysis would be a useful technique for the study of bonding characteristics between electrode materials and quartz surface, gold-quartz diffusion studies, and migration behavior of various electrode platings under shock or other extreme environmental conditions. Cathodoluminescence offers interesting possibilities for the study of chemical bonding under a variety of experimental parameters.

The authors wish to thank W. H. Hicklin for his help and interest in this work.

References

1. T. E. Everhart, O. C. Wells and W. C. Oatley, J. Electron. Control, 1, 97 (1959).

2. J. R. Banbury and W. C. Nixon, Scanning Electron Microscope Symposium, IITRI, Chicago, 1970, p. 473.

3. R.S.G. Goosdoves, G. V. Spivak, A. E. Lukianov, M. V. Bicov, and G. V. Saparin, Proceedings of the Fourth National Conference on Electron Microprobe Analysis, Pasadena, California, 1969, p. 58.

4. R. J. Gerdes and C. E. Wagner, Appl. Phys. Letters 18, 39 (1971).

5. H. Fukuyo, Bull. Tokyo Inst. Technol. Ser. A. 1, 1955. Also, I. Koga, Y. Tsuzuki, S. N. Witt, Jr., V. K. Woodcox, R. B. Belser, and W. H. Hicklin, Signal Corps Contract No. DA-36-039 SC-78905, 1960.

6. G. Sauerbrey, Z. Physik 178, 457 (1964).

7. R. A. Young and C. E. Wagner, J. Appl. Phys. 37, 4070-4076, (1966).

8. C. E. Wagner and R. A. Young, J. Appl. Cryst. 2, 39, (1969).

9. T. E. Everhart, Symposium on the Scanning Electron Microscope, IITRI, Chicago, 1968, p. 1

10. N. C. McDonald, G. Y. Robinson, and R. W. White, J. Appl. Phys. 40, 4516 (1969).

QUARTZ CRYSTAL UNITS FOR HIGH G ENVIRONMENT

Marvin Bernstein
Electronic Components Laboratory
U. S. Army Electronics Command

Summary

The design of small quartz crystal units, operating at the fundamental mode frequency of 15 to 23 MHz, which will survive shock acceleration amplitudes of 15,000 g's and more is given. Test data on the number of failures and the permanent frequency change as a result of various levels of shock are shown. A new method of bonding the quartz resonator to the support, using electroplated nickel films with thickness of the order of 0.5 thousandths of an inch is discussed. Precautions in the processing of the quartz resonator are shown to be necessary to avoid destruction of the resonator at shock levels higher than 15,000 g's.

Introduction

Test shock amplitudes for military quartz crystal units have been restricted to values of the order of 50 to 100 g units because of the fragile nature of the resonator. Applications in small equipments now require a more rugged crystal with levels in the range of 500 to several thousand g's. In addition, equipment has been developed which must operate after a 15,000 g shock, and telemetry equipments used in high altitude research projects require crystals capable of operating after acceleration at greater than 50,000 g's.

Experimental crystal units of the type shown in Figure 1 have been assembled in the Electronic Command's Electronic Components Laboratory that have operated after being shocked at levels up to 29 thousand g's with durations of several milliseconds. The holder is a standard HC-25/U enclosure with two stainless steel supports of the type commonly used with this holder. This crystal unit differs from those usually made in this enclosure by the crystal industry only because the method of bonding is new and the resonator is somewhat smaller.

Relatively little development has been undertaken in the area of high g crystal units except for that reported by the Harry Diamond Laboratory[1] and a paper presented at the International Telemetering Conference (ITC).[2] The present paper documents the continuation of the development reported on at the ITC in October 1970.

The majority of the shock tests were conducted at the Ballistics Research Laboratory using a high g acceleration testing method developed there.[3] The test projectile containing the crystal units is propelled out of the gun shown in Figure 2. The gun is a 5 inch smooth bore unit and from four to six ounces of powder is used to accelerate the projectile, at less than 1000 g's, in the gun barrel.

The projectile is decelerated by impacting the lead targets shown in Figure 3. Each lead block weighs 315 pounds and the projectile penetrates several of the blocks before coming to rest. The shock amplitude can be varied by changing the amount of powder while changing the shape of the nose tip causes a variation in the time duration of the shock.

Crystal Unit Design

Figure 4 shows the construction of the crystal mount and method of bonding. A narrow strip of Cr-Au or Cr-Cu is evaporated along the edge of the crystal resonator and the support clips are pressed over this composite metallic layer. The Cr results in excellent adhesion to the quartz while the upper layer of pure Au or Cu is used to effect the electroplated bond to the supports. The frequency of the resonators used varied from 10 MHz to 23 MHz, fundamental mode, and the diameter was 0.200 inch.

The basic premise in the development of the high g crystals was to use a short stiff mount and a small diameter resonator. An initial experiment using conductive epoxy cement indicated that this material lacked the necessary mechanical stability and could not be adequately controlled to be useful in the fabrication of these crystals. Electroplated Cu and Au bonds had been proven to be effective in the fabrication of glass enclosed crystals in this laboratory for more than a year. Furthermore, it was known that a nickel sulfamate plating solution is commercially available with certain outstanding characteristics.[4] This solution can plate nickel films having very low mechanical stress and freedom from brittleness, cracking, or crazing. The first experimental nickel electroplated bonds were easily achieved and exceptionally strong. Pull tests showed strengths of several pounds and failure always was the result of crystal resonator fracture. All of the crystals subjected to the high g shock tests used the holder HC-25/U and the nickel electroplated bond except for the first units constructed in the TO-5 enclosure.

The stainless steel ribbon supports must be given a thin electroplated layer of Ni by means of a strike plating solution which is chemically different from the sulfamate nickel plating solution. A useful plating bath solution can be made with 32 ounces/gal. of Ni Cl_2 .6 H_2O and 1 pint of HCl. This acid solution is used at a bath voltage of 2-3 volts, which should result in a thin nickel film in 10 to 15 seconds of plating time. This film must be very thin since a thick film is stressed and will crack and peel from the supports. The solution is very corrosive and should be immediately washed from the holder parts in hot flowing water and the parts dried to protect the thin nickel flash.

After the nickel strike has been applied to the supports, the resonator with the Cr-Au evaporated stripes are placed into the clip mount. The type of supports used in all of the high g crystals has been the Kay subminiature clip mount since the design is such that the contact line can be located very close to the edge of the resonator. The evaporated bond stripes are approximately 0.25 mm in width and the clip is centered on the bonding stripe. It is important that the resonator not contact the vertical portion of the mount, by pushing the resonator fully into the clip, since it will bond to this part of the mount with a resulting strain induced into the resonator.

The nickel bond is achieved, using the nickel sulfamate solution as purchased, at room temperature with a bath voltage of the order of 1.2 to 2 volts. The anodes used in both the strike bath and the bonding bath were pure nickel; usually, 8 resonators were bonded at the same time. The thickness of the nickel bond is determined by the product of the current and bonding time used. All of the units assembled for the high g tests were bonded at 1½ ma per crystal unit for 90 minutes or 135 ma-minutes. With 8 crystal units, the total current was, of course, 12 milliamperes. The time period can be shortened, with an increase in current but a point will be reached where the plated nickel is rough and obviously of poor quality. The bond has a thickness of the order of 0.5 thousandths (12.5 microns) with the 135 ma-minutes of plating. One test which used somewhat thinner and thicker bonds was inconclusive in so far as a crystal failure was concerned. A detailed technical report will contain the results of an extensive series of tests in which the bath plating current and length of plating times were varied with the bond strength measured in each case.[5]

Two methods of applying the electrodes to the resonator have been used. Some of the crystals made had Cr-Au stripes applied by evaporation and the resonators were then mounted within the clip attachments. The nickel electroplating operation was completed and the mounted and bonded resonator cleaned. Electrodes were then applied to the resonator by evaporation of Au and the frequency adjusted to the desired value. A second method was used with good success and in this case the mounting stripes and the base plate electrode were applied by evaporation of Cr-Au. In this method it is important not to apply too thick a layer of Cr since the alloy of Cr and Au has a very large ohmic loss. Figure 5 shows that the center of the base plated electrode and part of the tab are masked off with a lacquer material. Common nail polish is very useful. After allowing this material to air dry for a few minutes, the electrobond is made in the normal manner. The dried nail polish is easily removed with acetone. The units can then be cleaned and adjusted to frequency by either using electroplated Ni or additional evaporated metal.

The visual inspection of the quartz resonator, using a 20 X microscope, is a necessary operation. Scratched surfaces, body fractures, or chipped edges are all reasons for rejecting the resonator. The defects in the quartz resonator can be made much more easily visible by etching the surface. The amount of etch required is greater than that specified in MIL-C-3098, however, to clearly show all of the surface scratches. In one experiment in which the resonator was repeatedly etched, the number of visible scratches continued to increase until the lapped resonator was almost clear and the frequency at 17 MHz had increased by 1 MHz. Even when extreme care is taken, however, some resonators will break when subjected to shocks with levels of 20 Kg's or more. Preliminary results of one test seem to indicate that if the resonator is immersed in a fluid in a high power ultrasonic cleaner (20 KHz and 150 watts), approximately 30% of the resonators break. The appearance of the breaks, either fairly straight cracks across the whole resonator or small fractures in the body of the resonator resemble the inoperative crystal after the high g shock. There is a possibility that this means could be used to expand submicroscopic flaws in the resonator and thus reduce the number of failures in the shock test. Crystal resonators that are fracture-free can be made to be potential failures if care is not taken in effecting the seal between the holder cover and base. It has been observed that crystal resonators frequently had small fractures leading out of the Ni bond and these fractures appeared to predominate at the point where the clips are bonded to the resonator. It is thought that this is due to the very good thermal coupling between the Ni bond and the resonator. When the holder is heated too rapidly, the quartz under the bond expands and causes the quartz outside the Ni bond area, which is cooler, to fracture because that quartz is now in tension. Allowing the bases to heat up slowly, in the order of 30 seconds to reach solder melting temperatures, has reduced the problem to some extent.

Test Results

A fairly large number of tests were conducted in which the crystal units were subjected to shocks of 10 to 29 Kg's. Figure 6 shows the results of a series of tests beginning with the first test of a 10 point mounted resonator in a TO-5 enclosure. This unit was bonded with conductive epoxy, and the very large frequency changes apparently were the result of mechanical changes in some of the bonds. After this initial test, all the remainder of the units used Ni bonds and a 2 point support. Except for the one test in an air gun at Sandia, all tests were conducted at Ballistics Research Laboratory in the high g lead test facility at shock levels of 15 to 29 Kg. Figure 7 shows additional data at the higher g levels. Some of the failures in group 9 were due to reworking of the units because of holder damage sustained in an earlier test. The principle cause of these failures was the initiation of small fractures in the edge of the resonator by the too rapid heating of the base when melting the solder seal. The data on the series of shock tests, with respect to permanent frequency change, is shown on Figure 8. The horizontal bar shows the average frequency change and the 1-sigma limits are defined by the ends of the vertical bars. There is an obvious increase in both average frequency change and the standard deviation with increasing amplitude. One 3rd overtone crystal unit, at 30 MHz, was subjected to a 22-Kg shock and changed frequency less than 1 part per million possibly indicating that overtone operation, with a resulting thicker resonator, should be used in an application requiring a smaller change in frequency.

The number of units which failed at levels to 15 Kg's was small, amounting to but 3%. The one failed unit was made at a time when the condition of the quartz resonator had not previously been inspected with the microscope. At shock levels between 20 Kg's and 29 Kg's, 17% of the units failed. This figure is somewhat high and is due in part to the inclusion of resonators with obvious defects to learn if in fact, an inspection could be used to eliminate potential defective resonators. When a fracture occurs in the central area of the resonator, the activity is poor and the unit is an electrical reject. Some fractures, however, can occur in the outer edges of the resonator and these crystals can be electrically acceptable but will always be destroyed in the shock test. Figure 9 shows the result of a fairly deep curving scratch across the resonator which apparently caused the low power level resistance to greatly increase. The possibility exists that this test might be useful to separate potentially defective units since these units occasionally fail at shock levels higher than 15 Kg's. It is important to round the resonator carefully to avoid edge fractures; these will expand and allow edge chips to form resulting in surface scratches during the lapping step. Even more important, the edge fractures can expand into the quartz during the shock test thus causing failure.

The two-point mounted resonator has a preferred acceleration direction as shown in Figure 10. When accelerated in the direction away from the pins, the mount supports are in tension and the resonator has maximum stiffness and fewest percentage failures occurred in the shock tests. In one test of a group of units procured from a commercial source, by the Harry Diamond Laboratories, all units failed at 18 Kg's when accelerated in the direction marked "good".

In the same test, three out of twelve units made in the Components Laboratory failed. The principal difference in the commercial units was the use of a small bonding spot rather than the long stripe shown in this figure.

A few crystals were subjected to shocks in the plane in which the resonator flexes and these failed at 10 Kg's. There was evidence that the ribbon support welds were inadequate since the supports rotated to a degree that the resonators might have contacted the enclosure.

Two units with 10 MHz fundamental frequency were subjected to a 22-Kg shock in the direction in which the supports flex. Both support mounts bent severely and the crystals became detached from the supports but did not fracture. The Cr base film was stripped from the semi-polished resonator while on the lapped resonator, the Ni bond ruptured.

For applications where there is substantial rotational forces in addition to direct acceleration, it is necessary that the crystal unit be mounted as close to the axis of rotation as possible. As of the time of writing this report, no data has been obtained on the survival of these crystals in a test which included rotation. The gun used at BRL, as mentioned before, is a smooth bore unit so no rotational forces are generated.

Figure 11 shows a resonator with a crack commonly found in one group of resonators that had been processed with inadequate care. A lot of 100 rounded resonators were obtained from an outside source and inspection showed many with edge and body fractures and these units were discarded. The remainder were cleaned, etched and again inspected under a 20 X microscope and additional resonators with cause for rejection were found. Upon plating and mounting the good units, inspection again located defective resonators. The continuing loss of units occurred at test before sealing the enclosure and after the seal was accomplished. Out of the initial group of 100 resonators, only twenty-two units were electrically acceptable for the shock tests. At 15,000 g's, three out of twelve units failed while at 29,000 g's, nine out of ten units failed. From this experience, it appears that possible shock test failure cannot be eliminated by visual inspection of the quartz resonator.

Conclusion

The design of a low cost crystal unit, capable of surviving peak acceleration shocks of 15,000 g's at 2½ milliseconds has been completed. These units, when carefully fabricated, show catastrophic failure rates of 3% and average frequency changes of 5 parts per million. The electroplated nickel method of bonding the resonator to the mount supports has proven to be practical, precisely controllable and extremely reliable. Further work in the areas of a more rugged mount, better resonator inspection methods and preseal stress testing of the crystal must be undertaken to improve the performance at shock levels of 30,000 g's.

Figure 12 shows a three point mount which results in increased crystal resonator stiffness and an initial test indicates improved shock performance. It appears to be feasible to develop a crystal design in Holder HC-25, perhaps using 3rd overtone units at 20 MHz, capable of operation after being shocked at 50,000 g's.

Acknowledgements

The author wishes to acknowledge the enthusiastic help of Mr. Victor Richard, of the Ballistics Research Laboratory, for providing the use of the high g lead test facilities and to Mr. Tom Liss, of the Harry Diamond Laboratory, for additional high g tests. The help of H. Wasshausen, A. Dagavarian, H. Spaights and D. Boyce, of the Crystal Lab, was found to be indispensable.

References

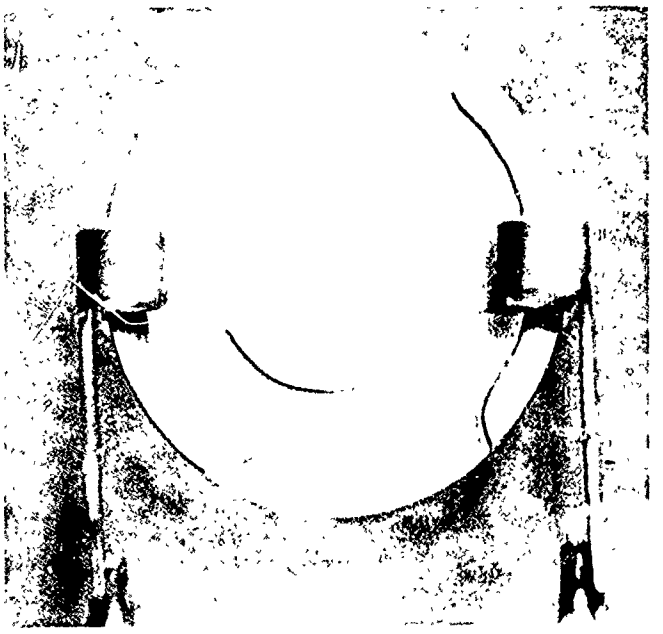
1. F. T. Liss, J. F. Richardson, "Ruggedized Quartz Oscillator Crystals for Gun-Launched Vehicles," TM-68-23, Harry Diamond Laboratories, Washington, D. C. 20438.

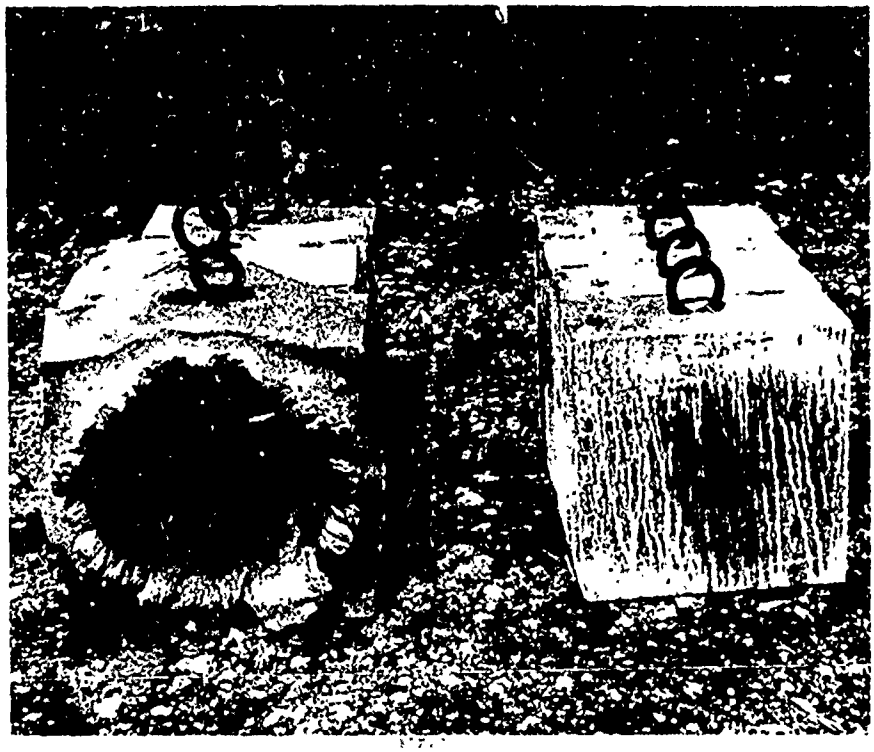
2. 1970 International Telemetering Conference, International Foundation for Telemetering, 19730 Ventura Blvd., Suite 7, Woodland Hills, California 91364.

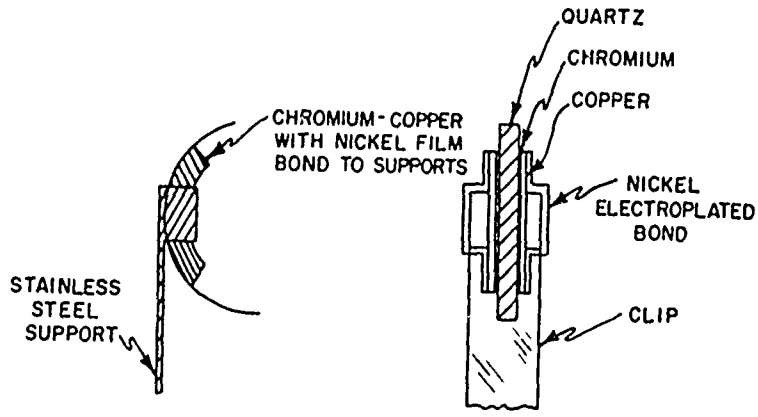
3. S. T. Marks, etal, "The Development of a High Acceleration Testing Technique for the Electronic Instrumentation of Harp Projectile Systems," Memo Report No. 1738, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland 21005.

4. Type SN, Sulfamate Nickel Plating Solution, Allied Research Products, Inc., Baltimore, Maryland 21205.

5. J. M. Stanley, "Nickel Electrodeposits," Technical Report dated July 1971. Electronic Components Laboratory, USAECOM, Fort Monmouth, N. J. 07703.







CRYSTAL CONSTRUCTION

FIGURE -

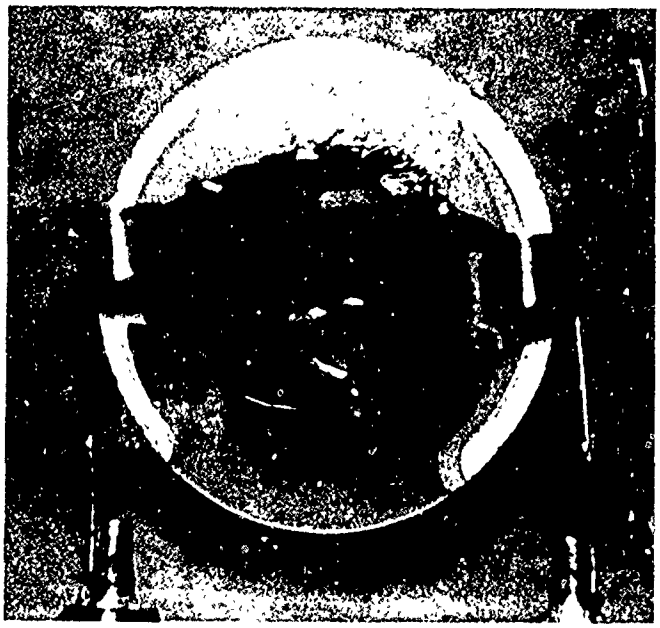


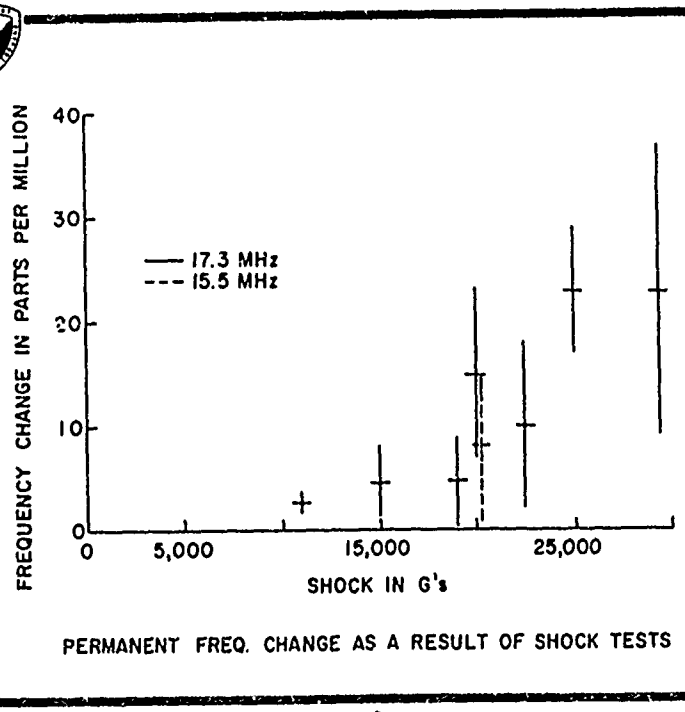
FIGURE 5

	TEST NUMBER AND HOLDER TYPE					
	1 TO-5	2 HC-25	3 HC-25	4 HC-25	5 HC-25	6 HC-25
TEST CONDITIONS	A 15,000 G SHOCK B. 20,000 G CENTRIFUGE	A 15,000 G SHOCK B 20,000 G CENTRIFUGE	A 15,000 G SHOCK	A 15,000 G SHOCK	A 10,800 G 3 M SEC (SANDIA)	A 18,500 G BRL
TEST RESULTS FREQUENCY CHANGE IN PARTS PER MILLION (PPM)	A 48 ppm AV 76 ppm 1σ B 11 ppm AV 10 ppm 1σ NO FAILURES	A 4 ppm AV 3.7 ppm 1σ B 0.8 ppm AV 0.5 ppm 1σ 19 OUT OF 23 FAILURES HOLDERS DAMAGED IN TEST	A 8.4 ppm AV 2.9 ppm 1σ 10 TESTED NO FAILURES BASES DAMAGED IN TEST	A 4.6 ppm AV 3.5 ppm 1σ 8 TESTED 1 FAILURE	A 2.7 ppm AV 1.1 ppm 1σ 10 TESTED NO FAILURES 15.5 MHz	A. 4.6 ppm AV 4.5 ppm 1σ 13 TESTED NO FAILURES 17.2 MHz

FIGURE 6

	TEST NUMBER AND HOLDER TYPE				
	7 HC-25	8 HC-25	9 HC-25	10 HC-25	11 HC-25
TEST CONDITIONS	A. 22,000 G BRL	A. 25,000 G BRL	A. 20,000 G BRL	A. 20,000 G BRL	A. 29,000 G BRL
TEST RESULTS FREQUENCY CHANGE IN PARTS PER MILLION (PPM)	A. 10 ppm AV 8 ppm 1σ 10 TESTED NO FAILURES 17.2 MHz	A. 23 ppm AV 6 ppm 1σ 13 TESTED 3 FAILURES 17.2 MHz	A. 15 ppm AV 8 ppm 1σ 10 TESTED 4 FAILURES 17.2 MHz	A. 8 ppm AV 7 ppm 1σ 9 TESTED NO FAILURE 15.5 MHz	A. 23 ppm AV 14 ppm 1σ 10 TESTED 2 FAILURES 17.2 MHz

FIGURE 7



PERMANENT FREQ. CHANGE AS A RESULT OF SHOCK TESTS

FIGURE 8

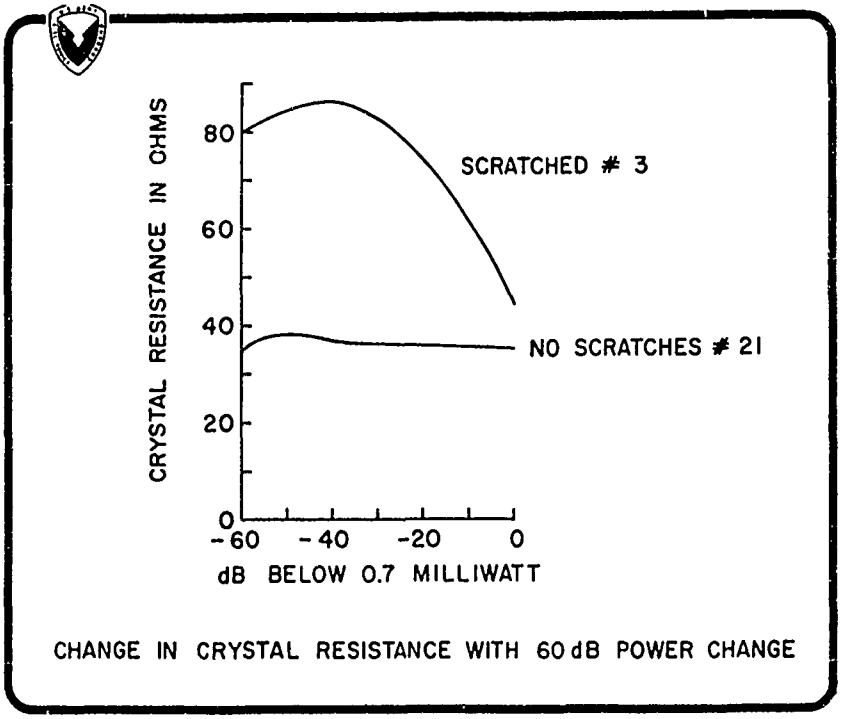


FIGURE 9

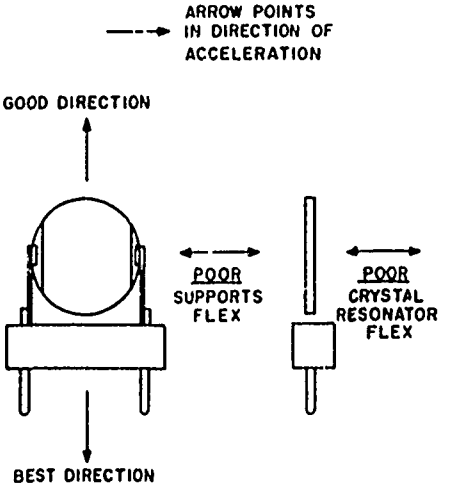
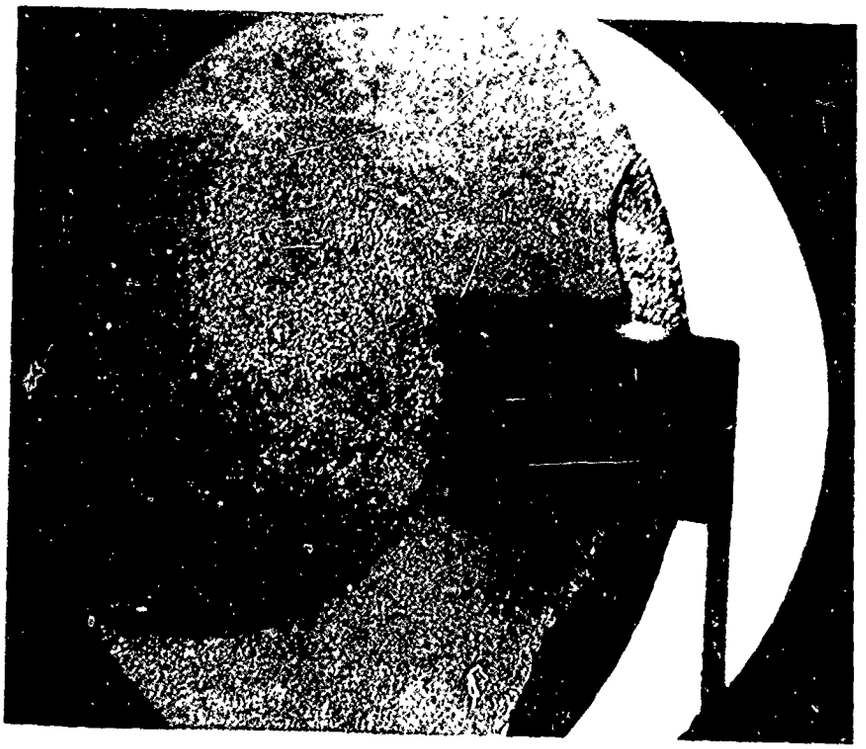
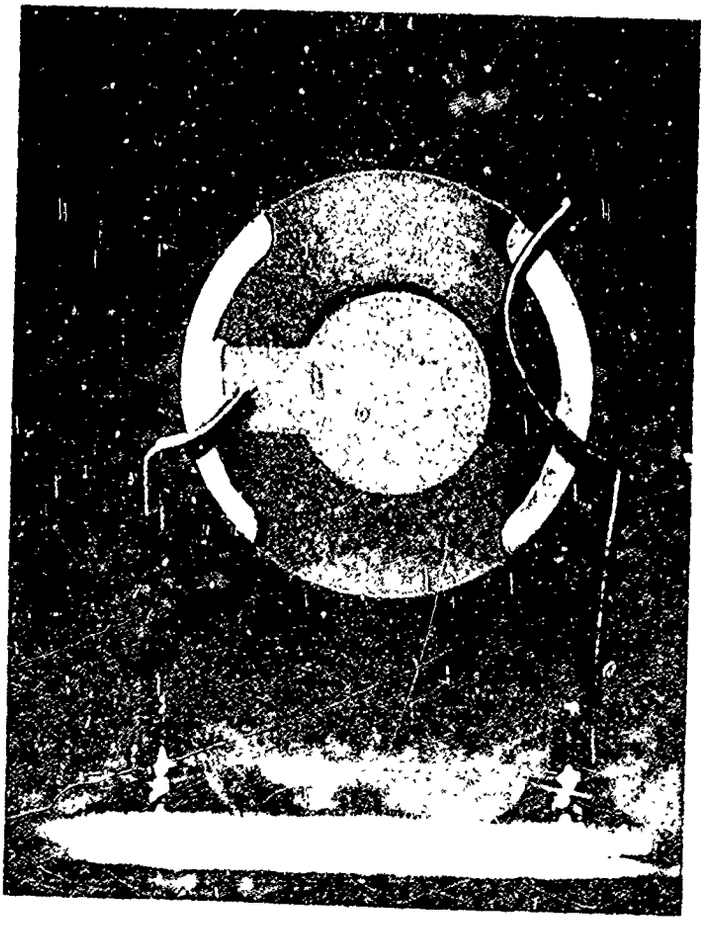


FIGURE 10





USING A PENDULUM DIFFRACTOMETER TO IMPROVE PRECISION OF X-RAYING QUARTZ CRYSTALS

Gerald E. Nemetz

General Electric Company

Lynchburg, Virginia

Summary

A new design pendulum suspension diffractometer for x-raying quartz crystals has been developed in the General Electric Mobile Radio Department. This diffractometer results in improved precision, higher productivity, and increased profitability in the production x-raying of quartz crystal blanks.

Introduction

In the 1950's the permissible shift in frequency of a radio transmitter was adequately controlled by ±3 minutes of tolerance on the ZZ' angle of an "AT" cut quartz crystal.

As the air waves became more crowded due to the rapid growth of the two-way radio industry, it became necessary for the Federal Communications Commission to apply more stringent controls. In the early 1960's the F.C.C. tightened the permissible transmitter shift in the ultra high frequency spectrum to ±2 parts per million. This ±2 parts per million tolerance for the transmitter translated to a ±½ minute tolerance on the ZZ' angle of the quartz crystal in the exciter stage. The double crystal goniometer would require a measuring system error not exceeding 10% of this tolerance, or ±3 seconds to be capable of satisfactorily measuring crystals to this tolerance.

Since the measuring system error of the goniometer exceeded ±3 seconds, the "temperature coefficient" test yields for "high stability" crystals (requiring ±2 parts per million tolerance) varied between 20% and 50% during 1964, 1965, and 1966. A number of projects were undertaken to improve the low "TC" yields. One of these was a project to improve the precision of the x-ray goniometer.

Problems Associated with the Goniometer

(1) Micrometer barrel is difficult to read due to the poor resolution. The smallest division is ¼ minute.

(2) Crystal work holding fixture locates the crystal on a flat surface. This requires the operator's wringing the crystal onto the flat surface (as two gage blocks are wrung together)--resulting in contamination of the work holder and blanks.

(3) Friction at the bearing shaft results in error in the ZZ' angle measurement.

(4) The x-ray shield requires two hand motions in x-raying one face, and four in the x-raying two faces of each blank.

Problems Related to the Goniometer Indicated by a Capability Study Performed in 1966

Process Control Engineering ran a process capability study FW-22 of 1966 on one x-ray goniometer station. During this period unfavorable statements had been made concerning the x-ray such as "Let's do away with x-ray"--we get the same "TC" yield with x-ray as we do without it." A look at the chart of the six operator's readings shows the basis for this statement. The angle readings of the six operators were widely dispersed. Note particularly that operator number 4 had no reading above the upper specification limit and 61% of the readings below the lower specification limit, whereas operator number 6 had 1% of the readings under the lower limit and 41% above the upper specification limit.

Pendulum Diffractometer

The pendulum diffractometer was developed in 1969. This instrument is constructed entirely of aluminum and stainless steel. The four table legs are constructed of heavy stainless steel tubing. The 28" X 36" X 2" thick table top is fabricated from a ½" stainless steel plate fastened on top of a 1½" thick aluminum plate. Two similar plates are mounted beneath the table top at a suitable foot rest height. Four levelling screws are provided for leveling the table.

The pendulum mount is bolted to the top of the table and consists of two carbide knife edges pointing upward. Mating with these knife edges are two carbide "V" shaped details brazed to a shaft which is free to rock through an angle. The pendulum arm is bolted to the shaft and hangs vertically downward. A micrometer head moves the pendulum arm through the "rocking angle" and a dial indicator provides the angle measurement reading directly in minutes and seconds of ZZ' angle.

An adjustable electronic transducer pickup is located in parallel with the dial indicator as an automation feature.

Capacitance Bridge - Proximity Shutter

The hand operated safety hood for cutting off the x-ray beam was replaced by the proximity shutter. The shutter is fabricated from a brass bar with a milled slot which allows the x-ray beam to pass through. The brass shutter is attached to an electrical d.c. solenoid which is actuated by the capacitance bridge. The sensor for the capacitance bridge is mounted near the crystal work holding fixture.

The brass shutter and solenoid assembly are mounted so that the x-ray beam passes through the milled slot in the brass shutter. As the operator brings her hand held work crystal to the work holding fixture, her hand is sensed by the sensor which actuates the solenoid - shutter assembly, moving the brass shutter in the path of the x-ray beam.

The proximity shutter provides increased productivity as it eliminates the hand motions required to raise and lower the metal hood.

Dial Indicator and Sorting Wheel

Several years ago I made a trip to an equipment manufacturer's plant to look at a diffractometer for quartz crystals which was offered for sale. This instrument used a large micrometer head for moving the arm through the "rocking angle". I asked the builder of the instrument how an operator would read the measured ZZ' angle. His answer was "You read the micrometer and then refer to a calibration chart".

A calibration chart may be acceptable for laboratory studies, but it would not be suitable for production x-raying due to:

(1) operator fatigue from referring to chart
(2) large potential error in cross-referencing
(3) slow method

A four inch diameter dial indicator which has .010 per dial revolution was selected, and the length of the pendulum arm was calculated to provide an angle measurement of five minutes per revolution of the dial indicator. A special scale which reads directly in minutes and seconds, with one second per division and five minutes per revolution was designed for use with the dial indicator.

A 10 segment color coded ring was attached to the bezel of the dial indicator. The ZZ' angle increment of each segment is equivalent to ½ minute.

The operator sorts crystal blanks into ½ minute increment of ZZ' angle by merely placing the blank into the jar with the number corresponding to the number on the ring indicated by the dial indicator pointer.

Under this sorting system the operator provides an objective sort since all ZZ' angle increments are indicated as acceptable.

Constrast this sorting method to the one whereby the operator sorts the acceptable blanks into one box, and the angle low and angle high blanks into their respective boxes. There are a number of psychological factors that influence the operator to place the blanks slightly outside the specification limits into the acceptable box.

Results

(1) A capability study comparing the pendulum diffractometer to the goniometer revealed the following:

   (a) The pendulum diffractometer has 1/3 the measuring system error of the goniometer
   (b) The intensity curve for the pendulum diffractometer is sharper and more symmetrical
   (c) The pendulum diffractometer can be peaked by turning the micrometer head either clockwise or counter-clockwise with less than a ±2 second difference from the initial ZZ' angle reading.

(2) The repeatability in readings of ZZ' angle of the pendulum diffractometer is such that 95% of the readings will repeat within ±2 seconds, providing the blank is not removed from the work holder.

(3) The improved precision of measurement in the pendulum diffractometer allowed for better correlation between ZZ' angle measurement and "TC" test data. The result is a 100% improvement in the "TC" test yields.

(4) The work station improvements due to the sorting wheel, proximity shutter, and MTS study resulted in an average increase in productivity of 330%.

Conclusions

	OLD RATE BEFORE IMPROVEMENTS	PRESENT RATE WITH IMPROVEMENTS	EXPECTED RATE WITH SORTING WHEEL
ONE FACE	100%	172%	296%
TWO FACE	50%	93%	189%

- AVERAGE INCREASE IN PRODUCTIVITY 330%
- AVERAGE INCREASE IN "TC" YIELD 100%

- INCREASED PRODUCTIVITY AND IMPROVED PRECISION RESULT IN:

HIGHER PROFITABILITY

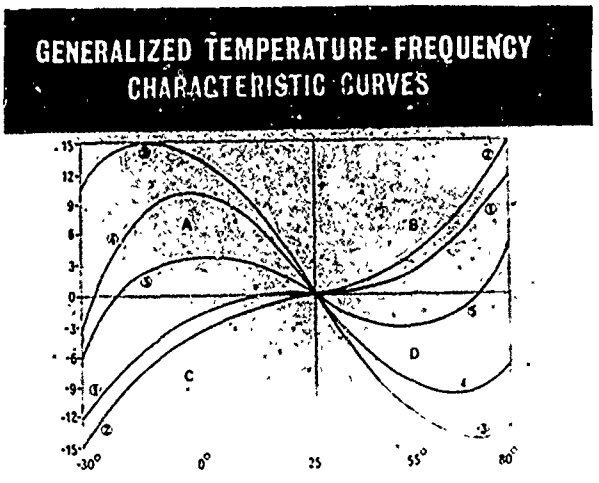


FIGURE 1 - TEMPERATURE COEFFICIENT CURVES

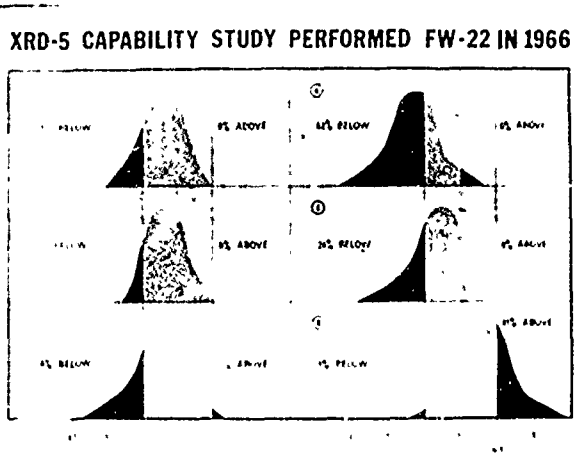


FIGURE 2 - GONIOMETER CAPABILITY STUDY

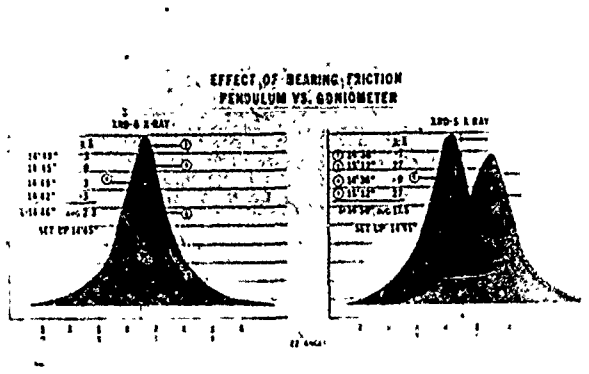


FIGURE 3 - EFFECT OF BEARING FRICTION

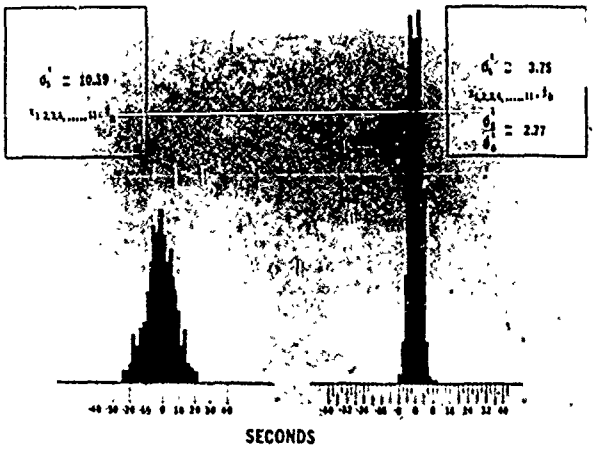


FIGURE 4 - COMPARISON OF XRD-5 GONIOMETER AND XRD-6 PENDULUM DIFFRACTOMETER

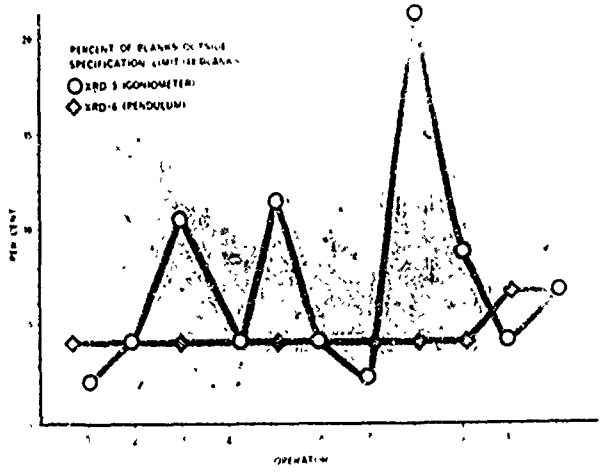


FIGURE 5 - COMPARISON - GONIOMETER AND PENDULUM DIFFRACTOMETER - CAPABILITY STUDY

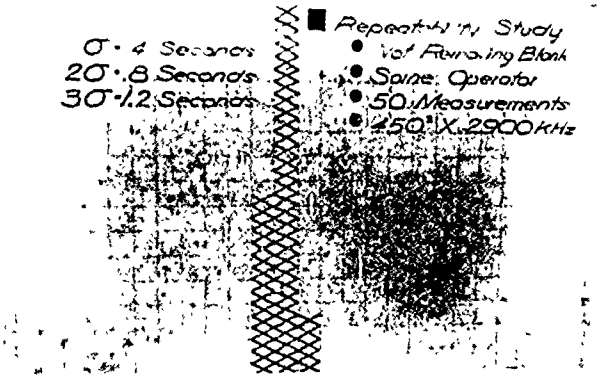


FIGURE 6 - PENDULUM DIFFRACTOMETER CAPABILITY STUDY

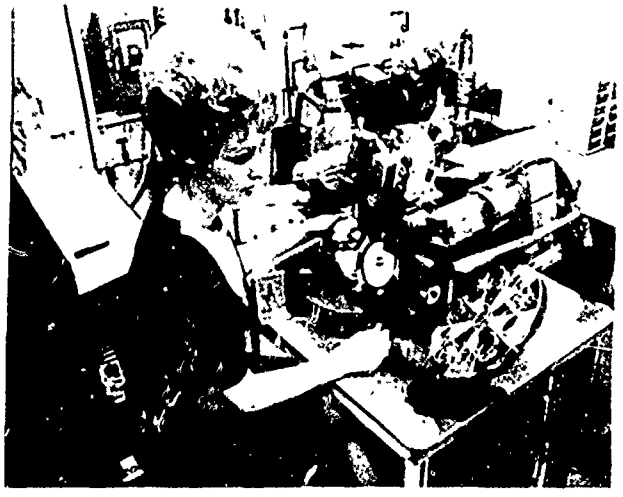


FIGURE 7 - PENDULUM DIFFRACTOMETER - FRONT VIEW

FIGURE 8 - PENDULUM DIFFRACTOMETER - SIDE VIEW

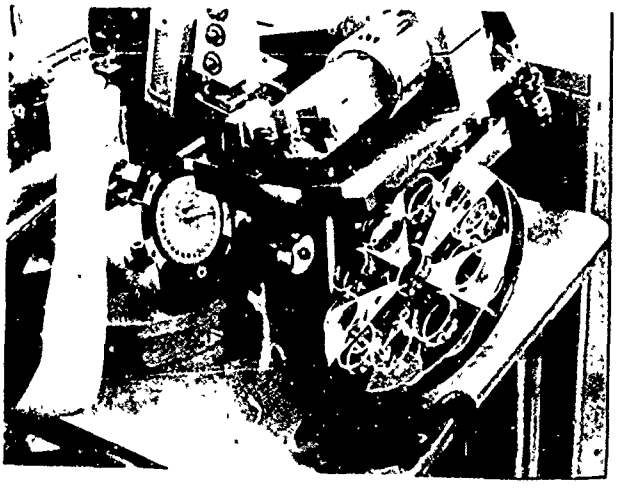


FIGURE 9 - PENDULUM DIFFRACTOMETER - OPERATOR LOADING BLANK ONTO FIXTURE

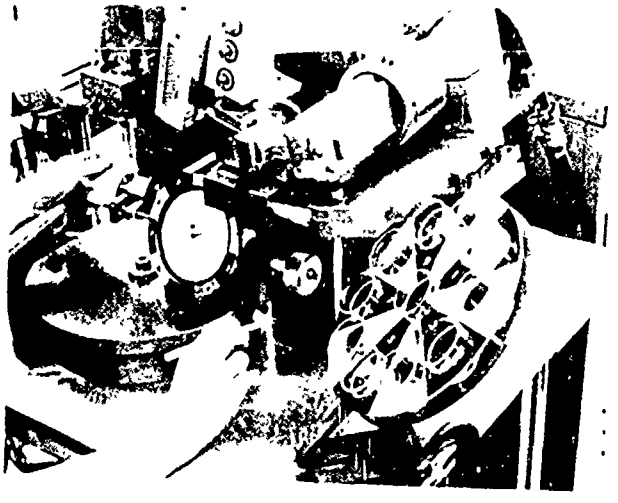


FIGURE 10 - OPERATOR ADJUSTING MICROMETER HEAD WITH LEFT HAND

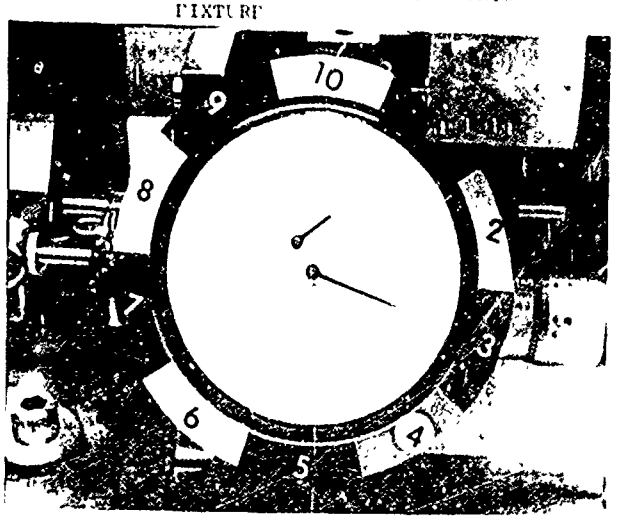


FIGURE 11 - DIAL INDICATOR

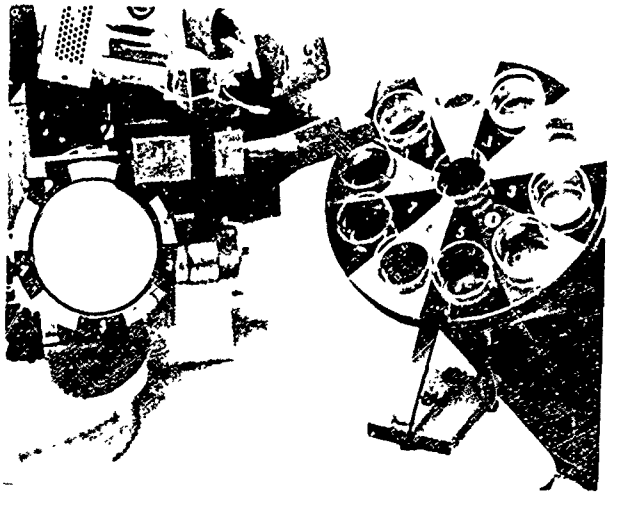


FIGURE 12 - DIAL INDICATOR AND SOLDER WHEEL

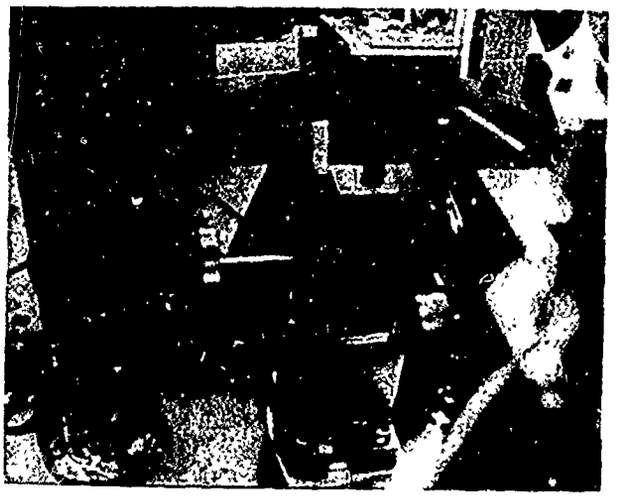


FIGURE 13 - KNIFE EDGE PENDULUM SUPPORT

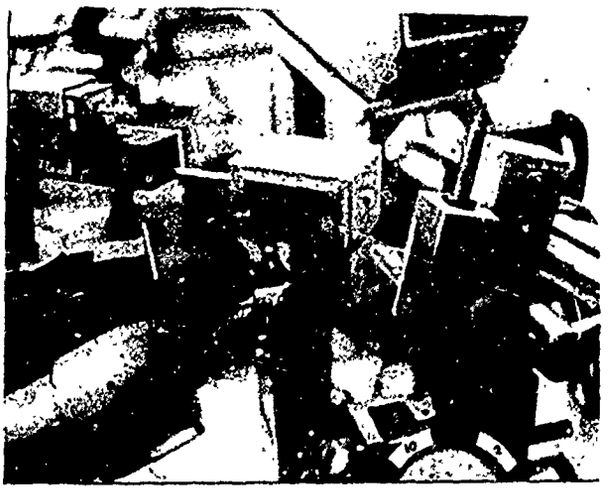


FIGURE 14 - CAPACITANCE BRIDGE AND SHUTTER ASSEMBLY

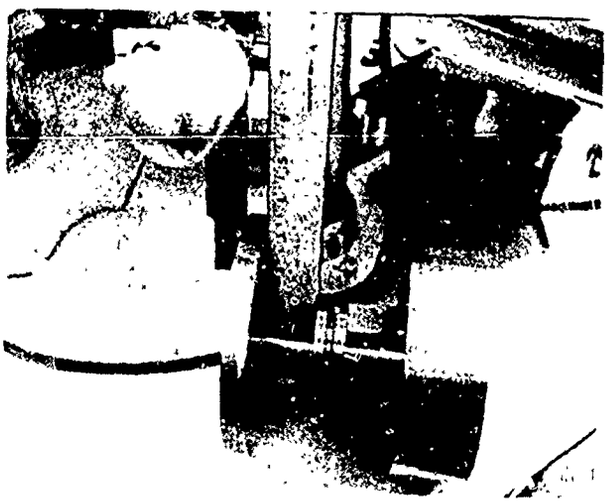


FIGURE 15 - SHAFT WITH CARBIDE "V"

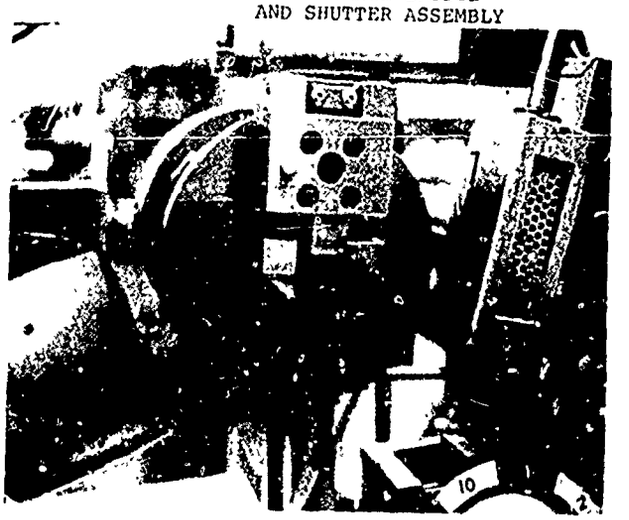


FIGURE 16 - THREE POINT CRYSTAL WORK HOLDING FIXTURE

THE CURRENT DEPENDENCY OF CRYSTAL UNIT RESISTANCE AT LOW DRIVE LEVEL

Shunsuke NONAKA, Tasuku YUUKI and Koich HARA

Transmission Division
Nippon Electric Company
Shimonumabe, Kawasaki, Japan

Summary

M. Bernstein described that, there are certain kinds of AT cut quartz resonator units that their resistance at such small currents as oscillator noise level are larger than that observed at normal operating crystal current. This large crystal unit resistance prevents the starting of oscillations. This phenomenon, that is, the large crystal unit resistance at such small current as oscillator noise level, is eliminated by overdriving the crystal. But, the resistance is gradually increased as the time elapses. This phenomenon was observed by help of an X-Y recorder and TS-683/TSM test set.

For measuring the crystal unit resistance, we didn't use such active circuits as TS-683/TSM test set, but we used such passive circuits using the transmission method.

As transmission method enabled us to measure crystal unit resistance over a very wide range of crystal current, ranging from very low levels such as noise level of the crystal oscillator, we could observe more quantitatively the behavior of the crystal unit resistance at those small current. Crystal unit resistance has maximum value at certain small currents, when there are such inhomogeneity or contamination on the surface of a quartz crystal plate such as scratches made while lapping or polishing, small metal particles. (For example, the maximum value is seven times larger than the resistance at larger current in a crystal controlled oscillator at normal crystal current.) The Resistance VS. Current Characteristics of the crystal unit is not reproducible before being overdriven, but it depends largely on its form of overdrive and its temperature. Current dependency of crystal unit resistance is completely eliminated by overdriving the crystal over the current given by

$$i_o = \frac{1}{100} \cdot \frac{f\gamma^2 \cdot d_o^3}{\sqrt{\gamma}}$$

where, i_o : current in mA
$f\gamma$: resonant frequency in MHz
d_o : diameter of active surface area in mm
γ : capacitance ratio.

Small metal particles sticking on the evaporated electrodes of the crystal were eliminated by overdriving the crystal over the above-mentioned currents. Thus, the current dependency of resistance caused by these particles were eliminated at the same time. The current dependency dit not appear upon storage.

To explain increased resistance phenomena we introduce the so-called static friction. When the displacement is small, the friction does not affect the movement of the crystal. When the displacement grows to a threshold amplitude, the friction suppresses the movement of the crystal. Suddenly, as the static friction is independent of the amplitude over the threshold, the effect of friction becomes smaller as amplitude grows larger.

We could explain increased resistance phenomena qualitatively. But, at this stage, we are not able to treat the problem quantitatively. Also, we cannot explain the problem concerning the hysteresis of the crystal unit resistance.

Introduction

To the best of our knowledge, there are very few reports on nonlinearity of crystal unit resistance, such as those by E.A. Gerber[1] and M. Bernstein[2], I. KOGA[3] and Ours[4],[5].

By measuring crystal unit resistance with active circuits such as TS-683/TSM test set, M. Bernstein showed that a combination of a thin contaminating film of oil and small particles of quartz increases crystal unit resistance at oscillator noise level.

The authors also discovered that crystal unit resistance increases at such low drive level as oscillator noise level compared with that observed at normal operating level. This is caused by small metal particles sticking on active crystal surface. The dependency of the resistance on the current, caused by small metal particles, is eliminated by overdriving. However, if the phenomenon concerning the overdriving of the crystal is not made clear, there will be still some probability of resistance dependency on the current, and the oscillation will not start.

In this paper the examples of current dependency of crystal unit resistance caused by various factors, and the effect of overdriving are described. Then, the mathematical formulation of current dependency of crystal unit resistance will be presented.

Experimental Procedure

Sample Preparation

Structures and frequencies of crystal units used in this experiment are shown in Table 1. A crystal plate has gold evaporated electrodes. In the case that it was desired to avoid the effects of evaporated electrodes, the crystal was held between the two concave metal electrodes as shown in table 1. In both cases the crystal plates were held in vacuum containers.

Measuring System

The resistance and the resonant frequency of the crystal unit were measured by transmission method using the passive circuit. We did not use the methods using the active circuits such as CI meter. The reason we chose the transmission method is that, it enables us to

measure these quantities over a very wide range of crystal drive current. The block diagram of the measuring system is shown in Fig. 1.

Experimental Results and Discussions

Typical Current Dependency of Crystal Unit Resistance

The IRE notations of the equivalent circuit of crystal unit are used in this paper. Besides, R_{1O} is defined to be the minimum value of R_1 in the range of measured current.

The typical data of current dependency of crystal unit resistance and resonant frequency are shown in Fig. 2. The resonant frequency of a crystal unit is given by equation (1) in terms of motional arm series resonant frequency of it.

$$f_\gamma = f_s \left(1 + \frac{1}{2} \cdot \frac{\gamma}{Q^2}\right) \qquad (1)$$

, where, $f_s = \dfrac{1}{2\pi\sqrt{L_1 C_1}}$, $\gamma = \dfrac{C_o}{C_1}$, $Q = \dfrac{2\pi f_s L_1}{R_1}$.

If f_s is assumed to be constant at any drive current, the maximum change in f_γ is of the order of 10^{-8}. Also, since f_γ increases as R_1 increases, f_γ is supposed to jump at the drive currents at which R_1 jumps. But the data of Fig. 2 do not agree with these two points. Therefore, Fig. 2 shows that both the crystal unit resistance and its motional arm series resonant frequency change with change in the drive current.

This paper is intended to give a clear picture of the mechanism that is thought to be directly responsible for increase of the crystal unit resistance.

Causes and Phenomenon of Increase of the Crystal Unit Resistance

Crystal unit resistance characteristics like that shown in Fig. 2 does not necessarily prevent the starting of the oscillations. Because, the increase of resistance is not so high. But, at low currents, the resistance increases too much, and the oscillation will not build up in the oscillator with the crystal unit. Fig. 3. In the case of the crystal unit whose resistance characteristics has amplitude suppressing effect as shown in Fig. 4, the oscillation builds up at a lower level than the normal operating level, and keeps oscillating with a stable amplitude.

Since these characteristics are easily affected by the previous forms of overdrive and temperature, there are various non-reproducible patterns of current dependency of the crystal unit resistance. However, they can roughly be divided into two groups. In one group, crystal unit resistance increases monotonously as the current decreases. Fig. 3. In the other groups, crystal unit resistance reaches its maximum at a certain drive current within the range of the measured current as shown in Fig. 4. Fig. 5 is a curve showing relationship between the temperature and current dependency of crystal unit resistance. This shows that the drive current at which a crystal unit resistance reaches its maximum, decreases as the temperature increases. The current dependency of resistance at high temperature looks somewhat like that shown in Fig. 3.

It can be concluded that there is no essential difference between the two patterns regarding the process in which the current dependency of crystal unit resistance is formed.

Fig. 8 shows the typical current dependency of crystal unit resistances. In the former two cases there are inhomogeneities on the surfaces of their crystal plates such as small gold particles and scratches as shown in Fig. 6 and Fig. 7 respectively. In the latter three cases crystal plates are coated with homogeneous thin films. There is a difference in crystal unit resistance between the two groups: in the former, the change of resistance is larger, where, in the latter it is small.

Overdrive

The current dependency of crystal unit resistance due to small gold particles sticking on the crystal plate was eliminated by overdriving it as shown in Fig. 9. The crystal unit which was overdriven above the threshold current, hence, was improved in loss did not have such current dependency of resistance as it used to have. Photographs in Fig. 10 show the change which happened on the above-mentioned crystal plate: the gold particle sticking on the electrode was eliminated by overdrive and a hole can be found where it had been. Naturally, resistance of the crystal unit, which was overdriven, keeps its improved value for more than a year as shown in Fig. 11.

The threshold current of circular thickness shear resonator is given approximately as follow:

$$I_{Th} = K \cdot \frac{f_\gamma^2 \cdot d_o^3}{\sqrt{\gamma}} \quad \text{(Refer to Appendix)} \qquad (2)$$

, where, I_{Th} : Threshold current in mA
K : Constant obtained by experiment
f_γ : Resonant frequency in MHz
d_o : Diameter of active area in mm
γ : Capacitance ratio.

This equation is obtained on the basis that the small particles sticking on a crystal plate will be eliminated when the maximum surface strain of it is larger than a certain value. K depends upon the vibration mode and the adhesive force between a small particle and crystal plate. $K \simeq 1/100$ was obtained experimentally for various kinds of AT cut resonators. This fact supports the fact that the small particles sticking on the surface of a crystal plate will be eliminated when the maximum surface strain of it is larger than a certain value.

This equation shows that the threshold current is smaller at lower overtone orders. In addition to this fact, generally speaking, active surface area is larger for lower overtone orders. In order to improve the Q of a crystal unit, it is more effective to overdrive it at lower overtone frequencies. Fig. 12 confirms the above interpretation. Crystal unit resistance at 7th overtone frequency loses its current dependency by overdriving it at the same frequency. In the case of the 5th overtone frequency, the overdriven resistance is less than the 7th one, and with the same character.

Mathematical Formulation

It is necessary to make it clear how current dependency of crystal unit resistance is generated. Because, it will help us for stabilizing the resonant frequency, and also for guaranteeing the reliability of its operation. Mathematical formulation of the phenomena of small gold particle is done as the first step.

At above the current at which the crystal unit resistance reaches its maximum value, the product of resistance change from the one measured at normal operating current and drive current is approximately constant. That is, $(R_1 - R_{1O}) \cdot I = \text{const}$ or

$$R_1 = R_{10} + \frac{V_o}{I} U(I - I_o) \qquad (3)$$

, where, R_{10} : Crystal unit resistance observed at normal operating current
V_o : Constant counter electromotive force
I : Drive current
I_o : Current at which crystal resistance gets its maximal value.

$$U(I - I_o) = \begin{cases} 0 & , \quad I < I_o \\ 1 & , \quad I \geq I_o \end{cases}.$$

Fig. 13 shows the behavior of R_1 VS. drive current represented by equation (3). The strain (S) at the point where the gold particle is, increases and when it reaches the $\bar{S}o$ level, the pulsating force (f) is exerted on the crystal surface in the same direction as of the friction. This process is repeated in every cycle, and is represented as follow:

$$f = \begin{cases} Fo\,\delta(S - \bar{S}o) & , \quad \frac{\partial s}{\partial t} \geq 0 \\ 0 & , \quad \frac{\partial s}{\partial t} < 0 \end{cases} \qquad (4)$$

Calculating mean value of dissipated energy in a cycle, the equation similar to equation (3) can be obtained easily.

This shows that the current dependency of crystal unit resistance is caused by such pulsating force represented by equation (4).

It was shown that the materials which cause this nonlinear term are small particles of electrode metal, those of quartz, or abrasive and oil film sticking on active crystal surface. Besides, twin crystals and cracks existing locally seem to have the same characteristics in generating nonlinear term as these small particles.

The process responsible for generating of nonlinear term is not known clearly yet, but there are two probabilities: the dependency of crystal unit resistance may be the result of the nonlinear force which is generated in friction, collision or viscocity between these materials and the crystal surfaces, or it may be from the nonlinearity between stress and strain in the crystal plate due to these forces.

Conclusion

1. Small metal particles sticking on active quartz surfaces not only cause the current dependency of the crystal unit resistance but also they cause the current dependency of the other constants.

2. There are various patterns of current dependency of crystal unit resistance, but there is no essential difference between them.

3. Typical current dependency of crystal unit resistance can be approximately represented by

$$R_1 = R_{10} + \frac{V_o}{I} U(I - I_o).$$

4. Small metal particles sticking on the active surface of a quartz plate can be eliminated by a current above $i_o \simeq \frac{1}{100} \cdot \frac{f_Y{}^2 \cdot d_o{}^3}{\sqrt{r}}$. Current dependency of resistance in these overdriven crystal units cannot appear again upon storage.

Acknowledgement

We are deeply indebted to many persons for the help received from them, and specially wish to extend our sincere appreciation to Prof. H. Fukuyo of the Tokyo Institute of Technology and Prof. M. Onoe of Tokyo University for their constant guidance, Dr. I. Koga of the KDD Research and Development Laboratory and Mr. K. Jumonji of the Electrical Communication Laboratories, NTT for their advice and their discussion on this study. Also, let us thank Mr. Hongo, Assistant Mangor and the other members of our Crystal Group of Development Department, Transmission Division of The Nippon Electric Company, who, without their help it was not possible to write this paper. It is a pleasure to thank M. Jam for helping translation.

References

(1) E.A. Gerber — "VHF Crystal Grinding" Electronics Vol 27, No. 3 pp 101 - 103, March, 1954

(2) M. Bernstein — "Increased Crystal Unit Resistance at Oscillator noise levels." Proc. 21st Annual Frequency Control Symposium

(3) I. KOGA — "Anomalous Vibration (Activity Dips) in AT-Cut (R_1) Plates" The Transactions of IECE of Japan. Vol 52-A, No. 6, June, 1969.

(4) S. NONAKA et. al. — "Current Dependency of Crystal Resonance Impedance." Papers of Technical Group on Component Parts and Materials. CPM70-9 June, 1970 (in Japanese)

(5) S. NONAKA et. al. — "The Temperature Effects of Current Dependency of Crystal Unit Resistance" '71 National Convention Records of the Institute of Electronics and Communication Engineers of Japan. No. 213 (in Japanese)

Appendix

Notations used here are those shown in Fig. A-1. Besides, K_1, K_2, K_3, K_4 and K are constant. The displacement in direction (Crystallographic axis) is approximately given by,

$$U = U_o \cos^2\left(\frac{\pi r}{2r_o}\right) \cdot \sin\left\{\frac{(2n-1)\pi y}{2y_o} \quad \cos \omega t. \qquad (A\text{-}1)\right.$$

Stored kinetic power is given by

$$\frac{1}{T}\int_0^T dt \int_{-y_o/2}^{y_o/2} dy \int_0^{d_o/2} \frac{1}{2}\left(\frac{\partial u}{\partial t}\right)^2 \cdot 2\pi r\, dr = K_4(\omega u_o d_o)^2 y_o \qquad (A\text{-}2)$$

This value is equal to the electric power stored in inductance L_1.

$$\frac{1}{T}\int_0^T \frac{1}{2} L_1 i^2 \, dt = \frac{1}{2} L_1 i_o{}^2 \qquad (A\text{-}3)$$

, where, i_o: mean value of current.

Then $U_o = K_3 \frac{i_o}{fr d_o}\sqrt{\frac{L_1}{y_o}} = K_2 \frac{\sqrt{r}}{f r^2 {d_o}^2}\ i_o$ (A-4)

The maximum surface strain $|S_f|_{max}$ is given by,

$$|S_f|_{max} = K_1 \frac{\sqrt{r}}{f r^2 {d_o}^3}\ i_o \ . \quad \text{(A-5)}$$

If it is assumed that if the maximum surface strain exceeds a certain value, small particles are eliminated.

That is,

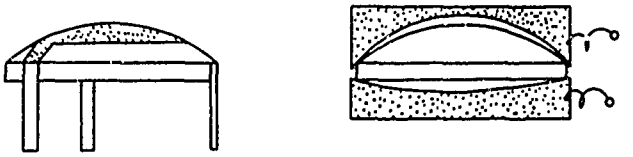
$$|S_f|_{max} = \text{const} \quad \text{(A-6)}$$

or $$i_o = K \cdot \frac{f r^2 {d_o}^3}{\sqrt{r}} \quad \text{(A-7)}$$

, where, K = const.

Frequency	Normal Operating Mode	Finish
2.2 MHz	Thickness Shear 3rd Overtone	Lapped
2.5 MHz	〃 5th 〃	Lapped or Polished
4.0 MHz	〃 3rd 〃	Lapped or Polished

Structure



Evaporated Electrode Type Gap Type

Table I Sample Crystals

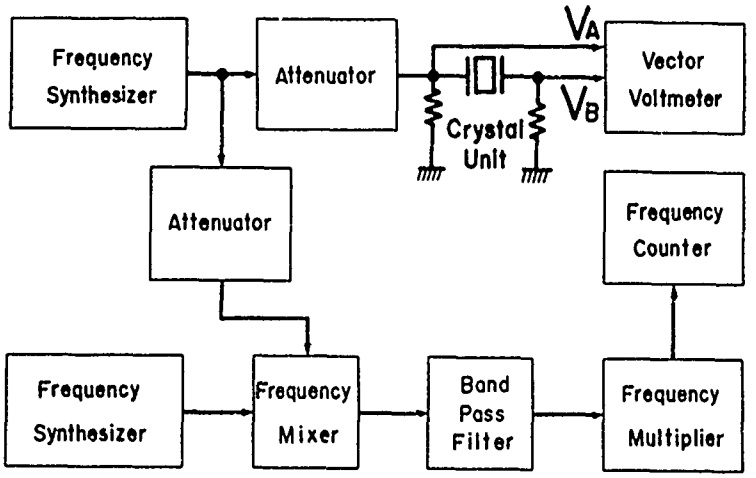


Fig.1 Measuring System

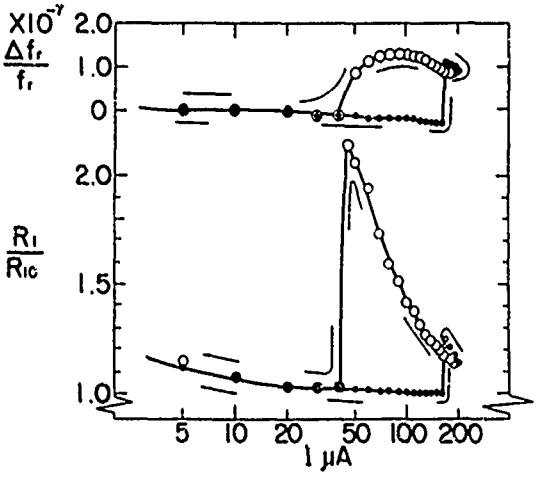


Fig.2 Typical Data of Current Dependency of Crystal Unit Resistance and Resonance Frequency

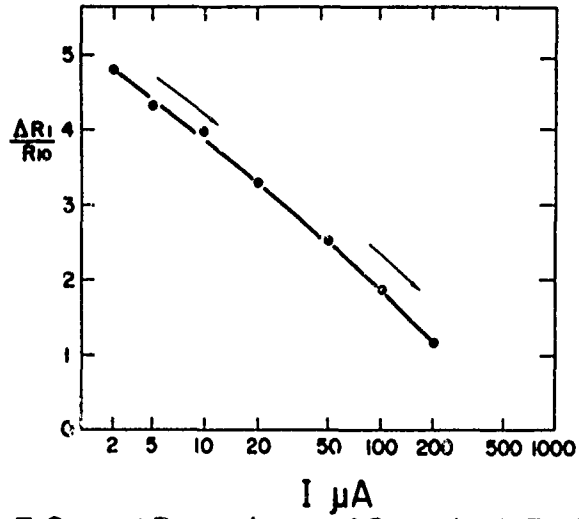


Fig.3 Current Dependency of Crystal unit Resistance

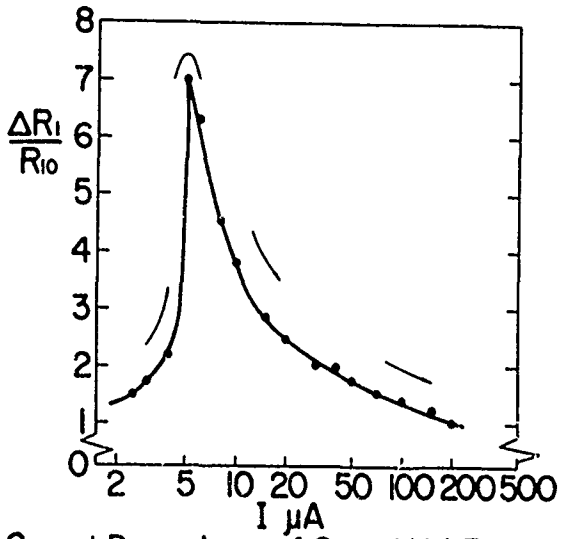


Fig.4 Current Dependency of Crystal Unit Resistance

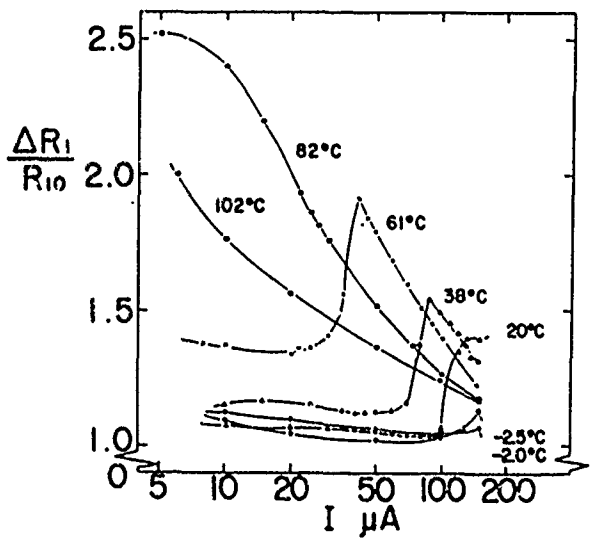


Fig.5 Relationship between Temperature and Current Dependency of Crystal Unit Resistance

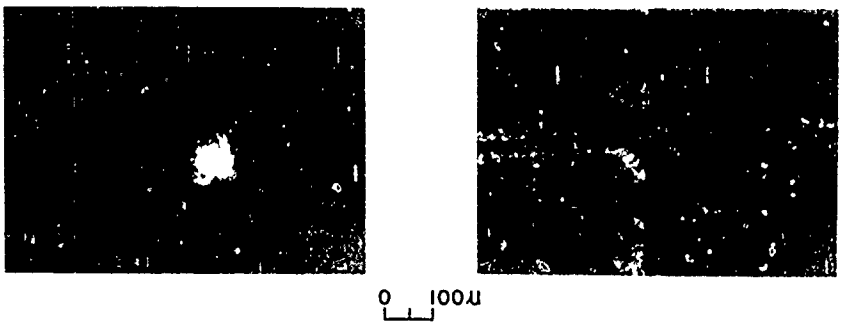


Fig.6 Gold Particle

Fig.7 Scratch

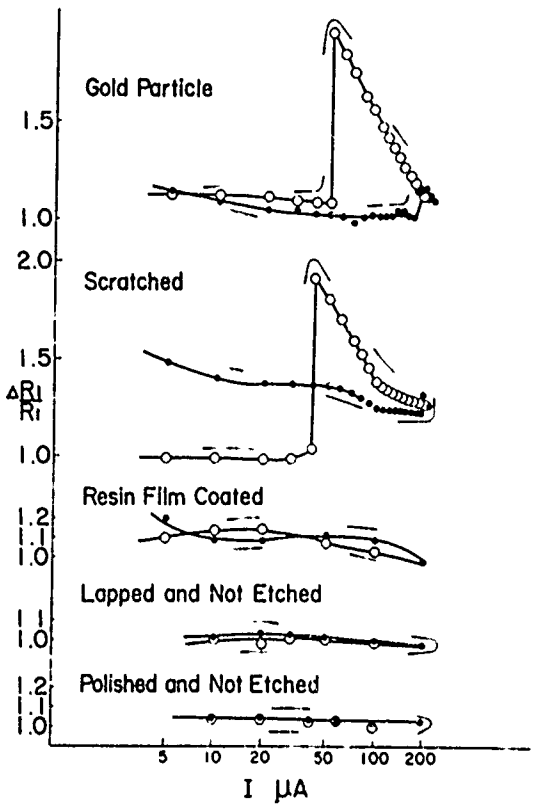


Fig.8 Typical Current Dependency of Crystal Unit Resistance Corresponding to Each Cause

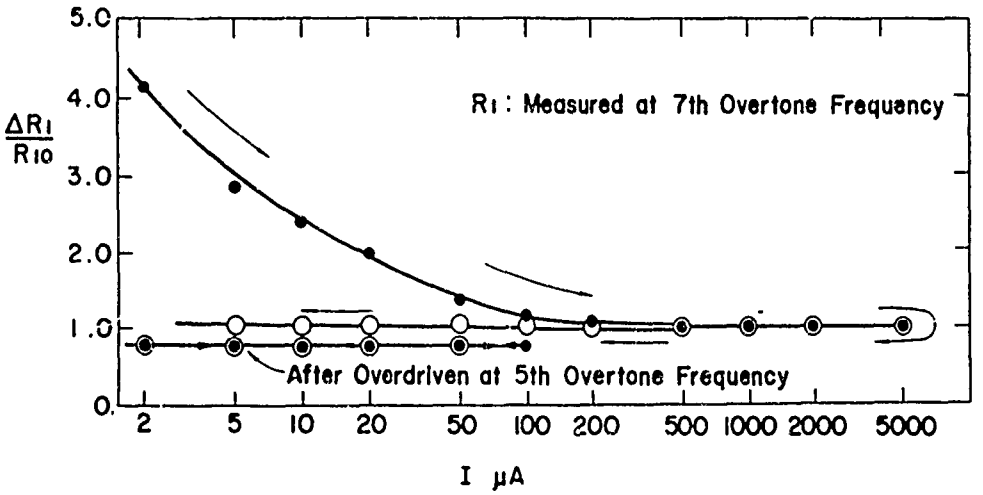


Fig. 9 Effect of overdrive

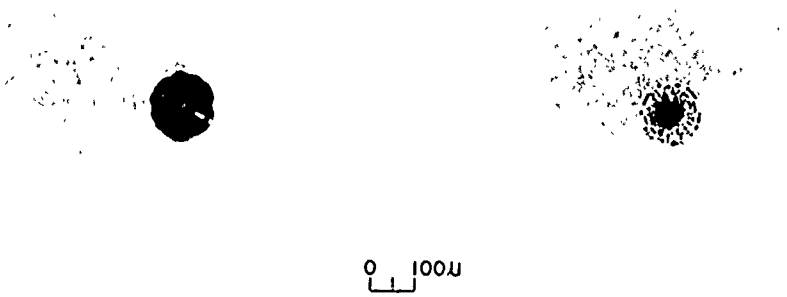


Fig.10 Change of Gold Particle by Overdrive

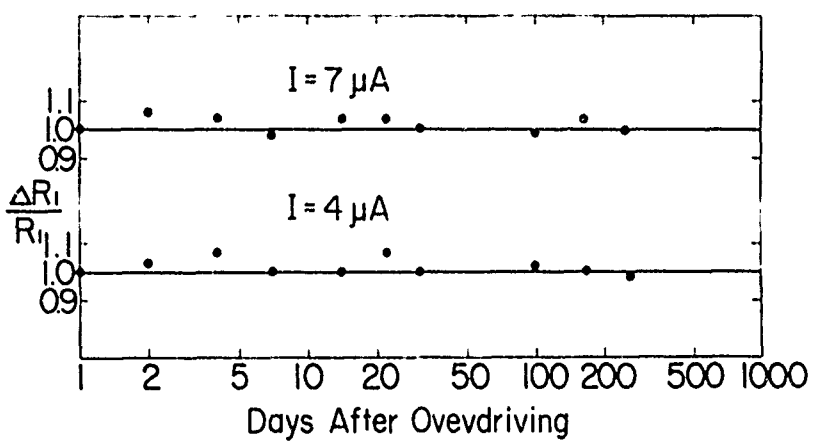


Fig.11 Resistance Ageing of Overdriven Crystal Unit

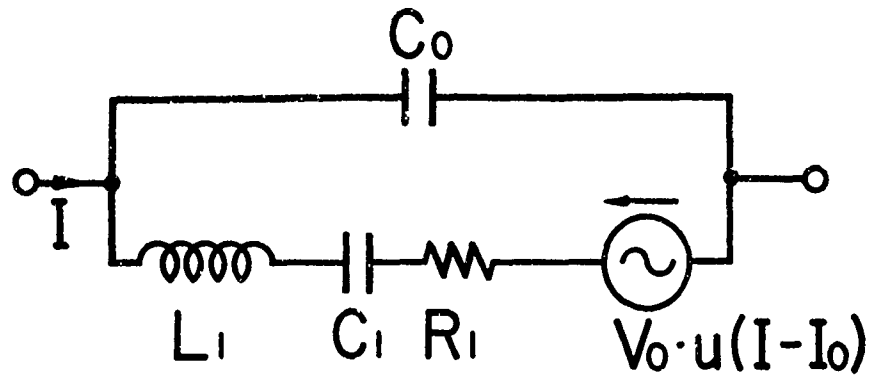


Fig. 12 Equivalent Circuit of Crystal Unit

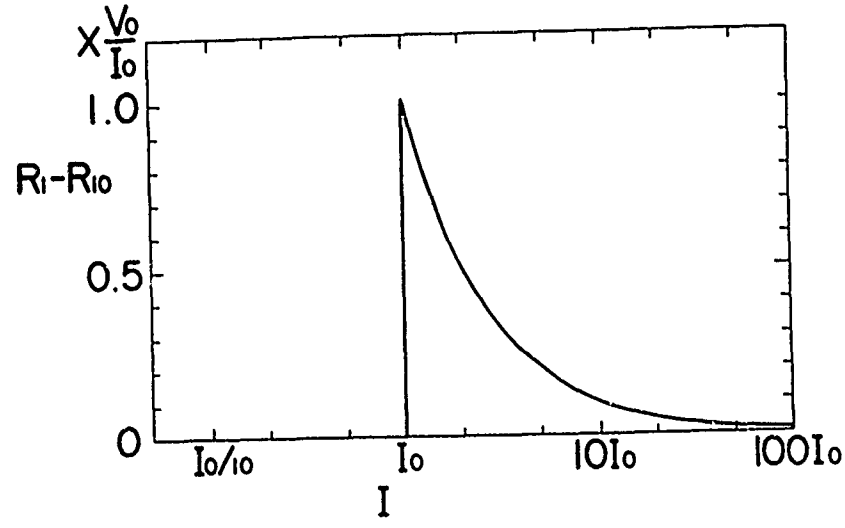
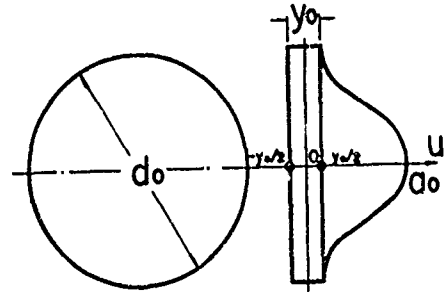


Fig.13 Current Dependency Calculated from Equation (4)



do : Diameter Active Area
yo : Thickness
uo : Maximum Displacement
n : Overtone Order
ρ : Density of Quartz

Fig.A -1 Structure of an AT Cut Crystal plate and its Displacement Distribution

STANDARDS AND THE FREQUENCY CONTROL INDUSTRY

John D. Holmbeck
Northern Engineering Laboratories, Inc.
Burlington, Wisconsin

Summary

There are many important reasons why standards should be formulated, understood, and used, in the frequency control industry. Unfortunately, in the past, their formulation has been left to someone else, they have been unnoticed or misunderstood, and too often either misused or unused. There has even been a fear that they will retard progress or damage a hard won competitive advantage. In practice, standards properly conceived and applied can reduce costs, improve the product and leave time and effort that might otherwise be lost "being original" to be expended on real progress. How much easier would it be to look up what already exists and would be suitable for the purpose, rather than starting from the beginning to design a new one. If we actively participate in standardization, it can provide a better forum for our point of view so that things can be done as we know they should be, rather than have designs and methods forced upon us by others with less knowledge of the subject and less interest in any single component, such as a crystal.

The United States is unique among developed nations in having a minimum of standards imposed by government. In dealing with other countries, we often find ourselves at a disadvantage because of lack of familiarity with their standards, but they are more familiar with ours than we are. This makes it relatively easy for them to export to us, but handicaps us in selling to them.

An explanation of the types of standards that exist on both national and international levels, and the mechanisms for updating them, should improve their utility and make participation more widespread. A standard should be a tool, rather than an obstacle.

Introduction

This paper will present an outline of how standards are developed, what they do (intentionally and unintentionally), and how they can be used to better advantage. Our industry is often forced to accept situations that are technically unsound simply because we do not have the respect of other disciplines and do not know what authority to cite to back up our positions.

History

The standards most widely used by us now came from the military in the 1950 era, and they have served a dual role reasonably well since then, but not without some disadvantages that can be overcome by action on our part. While the roles of military and commercial standards have some similarities, they also have some vital differences.

The history of WW2 communications will show that much civilian gear became the nucleous of the military systems at a tremendous saving of time, lives and money. It will also show that the chaos caused by the lack of crystal standards, either commercial or military, was a tragic handicap in the logistics of an explosively expanding need. Both areas can benefit greatly by timely action now when the technology is rapidly changing,giving us a chance to accommodate these changes without the mistakes of the past. One of the best ways to become familiar with standards and what they can do for us, is to participate actively in their formulation and revision.

As to the desirability of standards, let us look at a little history of our industry, and others, and then consider the future.

Prior to WW2 the United States crystal industry fell into two basic categories:

1. The large corporation with a crystal department set up primarily for supplying the needs of the parent company with little, if any, outside clientele.
2. The small independent producer who sold mostly on a "one of this or two of that" basis to anyone who could mail him between one and $100.00.

The latter group ranged from the one man, strictly custom operation, to the more sophisticated, a few of whom are still in the business today. Until the war made it necessary, there was practically no communication between these producers, and such standardization that existed consisted primarily of making crystal holders that would plug into tube sockets. The situation was not very much different in the rest of the world because crystals, outside the telephone field, were made for an infinite variety of applications and there seemed to be no need to standardize, because there was little volume production of a single item.

Before we got involved in the war, the United States, as the "arsenal of democracy", supplied equipment and parts that had been designed and proven in other countries, so we had to make crystals to fit their standards. (I use the word "standards" loosely in this case.) This was followed by our own needs when we began to re-arm. Since most pre WW2 U.S. military radio communication gear was not of sufficiently modern design to use crystals, as much was begged, borrowed, or stolen from our commercial and amateur fields as possible, with the result that literally hundreds of different enclosures, load capacitances and drive levels were used, where probably less than ten could have covered all the needs nicely. In addition, a whole new generation of gear was quickly developed, with little thought given to standards, adding to the proliferation.

Much of the "originality" came from not knowing what someone else was doing, a desire to have a unique product or a questionable "improvement". Due to the lack of idea exchange, technical capability varied, causing new types to spring up, simply because the manufacturer did not know any other way to make the item.

When the demands of the war suddenly required huge volumes of production, the government sought the guidance of the only high volume producer of crystals who, fortunately, also had the most concentrated scientific and engineering team in the field - the Bell Telephone System. While they did not do the whole job single-handed, they did have the necessary experience in both the technical and organizational phases of creating a large industry. With the aid of a few other groups and individuals, they provided such additional technology as was required to expand a few dozen small crystal shops into volume producers, as well as creating or converting a number of new ones.

Some of the first standardization in the industry came from the need for the crystal producer and equipment maker to agree on what was usable and what was not. This resulted in some "standard" test sets and test methods that were at least reasonably similar. Much of the technical information that was disseminated was merely to ensure that a part made in Kansas City would work in a set made in Chicago, which is really what standards are all about.

By 1945, a number of improved military crystal specs had been developed that served their immediate purpose fairly well, but they had a number of shortcomings. There was no similarity in format, so direct comparisons for choosing one over another were impractical. Test methods were "highly original" and other problems were plentiful. After a few years, the Armed Services Electrostandards Agency was formed by the military to bring order out of chaos. This resulted in the original version of MIL-C-3098, which again had much guidance from the Bell System.

For the first time we had standard formats, definitions and test methods; and crystals that were designed first so equipment could be designed around them, instead of the reverse. This was a system that the tube industry had used for many years with great success.

The reason that our first national standard was under military auspices was the fact that they were the only single customer served by practically every crystal manufacturer. Also, the volume and urgency of the requirements usually dictated that a single class of items should have more than one source.

The manufacturers of such components as tubes, resistors and capacitors, because their products are used in such large numbers, realized the necessity of standards as soon as the electronic boom began. This has worked to the advantage of everyone concerned because it not only helped to reduce the cost of producing the item, but also the cost of using it. Many new markets for electronic equipment that developed, would never have appeared if those industries had the lack of standards that prevail in our industry. Low prices don't mean loss of profit if the cost can go down faster than the price. Two things that bring cost down are volume and simplification of design and production. One of the greatest contributors to cost in the frequency control field, is specifications that are not really the same but neither are they different enough to justify two different specs. None of us, including the military, have clean hands on this score.

Let us take an example of a commercial product that represents some real volume crystal production - the 27 MHz Citizen Band market. How many different schemes are in production today? How many different load reactances, drive levels and holders? How much easier would it have been if a standard had been evolved by the crystal industry in which a single unit along the lines of the CR-23/U, had been proposed, along with a suggested circuit and some guidance, to the CB equipment manufacturer? How much better if each inquiry had been answered with a copy of the standard and with a more attractive price on the standard crystal than on a special one. Since that field is largely governed by price, the U.S. manufacturers could have saved money on some of the horrible examples of oscillator design, correlation troubles, specification changes and just plain mix-ups between customer and supplier, that occur today. There is no doubt that we could have at least delayed the takeover of a large part of that market by offshore producers. We might even have kept the production of the complete gear in this country longer if some of the stumbling efforts at oscillator design and the resulting crystal problems, with the added costs they create, had been avoided. When the HC-25 and HC-18 moved into the picture, along with transistors, we missed a second chance to do the job.

There are numerous other areas where we could have done more, or still can. The color TV field benefitted greatly by one manufacturer actually setting the standards in fact, if not in form. This has resulted in a minimum variety of specifications and has had a great influence on keeping most of the color TV crystal production in this country. In any field where we hope to develop a mass market, early standards can infl ence the choice of crystals over other devices, and by keeping costs down, can help to develop new markets.

Current Standards

The standards that our industry currently has in this country came from the following sources:

1. IEEE has provided the basic physics and engineering. These include:

   (a) Terms and definitions
   (b) Basic test methods

   Examples of these are:

   IRE #176, Standards on piezoelectric crystals.
   IRE #177, The piezoelectric vibrator: Definitions and methods of measurements. This is a document that every crystal manufacturer and user can find most useful.
   IRE #178, Determination of the elastic, piezoelectric and dielectric constants, the electromechanical coupling factor.
   IRE #179, Measurement of piezoelectric ceramics.
   IRE #180, Definitions of ferro-electric crystal terms.

   These standards are currently under review for updating.

2. EIA has provided manufacturing standards including:

   (a) Holders and Sockets

   RS-192-A, Holder outlines and pin connections for quartz crystal units. Work is constantly in progress to keep this standard up to date.
   RS-168-A, Dimensional and electrical characteristics defining tube and transistor sockets. (Includes crystal sockets.)

   (b) Production Tests

   RS-186-C, Standard test methods for electronic components.

   (c) Application Information

   Industrial Bulletin #6, Guide for the use of quartz crystal units for frequency control.

3. The military has the most used and complete U.S. National Standards, including many of the above factors, plus Article Sheets.

Most of these standards have been the result of the work of a few dedicated men, some of whom have served in two, or even all three, of these organizations, as well as the International Electrotechnical Commission, I.E.C.

Many other countries do not have exact counterparts to our three organizations as listed above, so they rely heavily on I.E.C. and use their documents as national standards. Due to the way our national documents come from different sources, the military has served as our single source of standards. This has led to much misuse and dilution of MIL-C-3098. The I.E.C., however, has one set of documents that includes the work of both IEEE and EIA. While the other I.E.C. countries have made some very valuable contributions to these documents, much of the information came from the United States. In many other countries, the I.E.C. standards are used the way we use our military specs in both military and commercial applications.

The I.E.C. has issued the following documents that can be very useful because they contain information that has come from so many sources and some of which is not readily available elsewhere in such convenient form.

Publication 122.1, Section 1, Standard Values and Conditions.
Section 2, Test Conditions.

By searching through IRE #177, EIA RS-186-C, and MIL-C-3098, you can find some, but not all, of this data. Some of it deals with definitions you will possibly find only on European specifications.

Publication 122.2, Guide to the use of Quartz Oscillator Crystals.

This is an international edition of EIA Industrial Bulletin #6.

Publication 122.3 This document contains the standard outlines and pin connections of EIA RS-192-A, as well as some information from RS-167-B, RS-168-A and MIL-C-3098.

Publication 283, Methods for the measurement of frequency and equivalent resistance of unwanted resonances of filter crystal units. I know of no other source for this particular information.

Publication 302, Standard definitions and methods of measurement for piezoelectric vibrators operating over the frequency range up to 30 MHz. This is actually an updated version of IRE #177. Work is now in progress to extend the frequency range upward.

Publication 314, Temperature control devices for quartz crystal units. This has no exact U.S. counterpart. It was written in Europe. Perhaps MIL-O-39021A comes as close as any, but is not as complete. The definitions, test methods and basing standards should make this document very useful.

The "Multipartite" agreement, which is forthcoming, will make it mandatory to comply with I.E.C. standards to ship components into the signatory countries. Most countries, with any kind of electronic industry, are expected to join the pact so the I.E.C. standards will be the "MIL specs of the export business".

TC-49, which is the I.E.C. Technical Committee on frequency control devices, is divided into a number of working groups. They are:

1. Crystals and Holders
2. Filters
3. Ovens
4. Ceramics
5. Cultured Quartz
6. Measurements
7. Oscillators

Most other countries, with their government owned and operated telephone and radio communication industries, are much more "national standard" oriented than we are. In order to live with all the standards, they had to learn to handle them reasonably well. As a consequence, while we have provided much of the technical leadership, their organizational experience has been most helpful.

With the alarm that was created by the Tripartite Agreement, which is now expanded to Multipartite, the United States government has decided to lend support through the Commerce Department to some of these activities, but it is not clear to what extent. Each industry that does not have its house in order can probably look forward to some government imposed "standards".

The Future

If we all do our part supporting standards, it does not have to be an expensive proposition, and it can be made to pay dividends. If we wait for the government to do it, every dollar of federal support will carry with it several dollars of federal control that can be an obstacle in domestic as well as foreign business.

One reason for the government involvement is that part of the Multipartite proposal is for a certification plan similar to our military QPL, for components to enter these countries. Some of them seem to feel that for such a certification plan to be of any value, the hand of national government has to be involved somewhere.

As a result of working with the various I.E.C. working groups, it became obvious that there would be considerable merit in having our national standards work organized in a compatible manner. Consequently, the EIA Engineering Working Group is now developing task groups to work in these various areas. The IEEE groups on Sonics and Ultrasonics, and Frequency and Time, also have appropriate subdivisions.

The ideal way for I.E.C. to work is to have each nation submit its national documents for comment and if the comments are favorable, they become the world standard. If we submit ours first, they become the standard. The comments, even if they are unfavorable, can help by showing up features that will make the product poor for the export market. Another advantage of leading, is it makes the less developed countries look to the United States for the equipment to update their electronic industry. It also helps to be able to discuss specs with export countries, in terms of I.E.C. standards, on type designations, dimensions, definitions, pin connections, and so forth.

One advantage of the I.E.C. documents over our National documents is the fact that a complete set of them can be obtained from one source. They include information that can be obtained partly from EIA and partly from IEEE, besides some foreign input that is not available from other sources. There is a need for a frequency control designers kit in this country that would gather all this information together between one set of covers and could be bought as a single item.

One reason the military specs have been so widely used and mis-used, is the fact that such complete coverage is provided in one document that can be ordered under one number and you don't get the bill till April 15th.

If these documents were used as they should be, the man hours now spent on negotiating technical details with customers, would pay for the time needed to comment on the standards to keep them correct and current.

There are a number of problem areas that need attention. In the past, we had a few key people who could always be depended upon to do what needed to be done when nobody else would do it. This certainly got things done, but it had a number of drawbacks. It left us with too much of a feeling that each of us did not have to do his part because somebody else would do it when it became urgent. It also put too many eggs in each basket. Because our standards were taken care of for us, we took them for granted and, in too many cases, never knew they existed, or at least what they provided for.

Both manufacturers and users of frequency control devices should take a good hard look at their attitudes and practices where standards are concerned. How often do you actually consult RS-192-A before you make a holder drawing? Do you ask for non-standard dimensions when standard ones would do just as well if you knew what they were? Do the words you use in your specs mean the same to you as they are defined in standard "Terms and Definitions"? How many engineers do you encounter who refer to "series mode" or "parallel mode" crystals as if they were as different as fundamental and overtone, indicating that they are in need of some homework on IRE #177, or at least EIA Bulletin #6?

There are a number of ways that you can help yourself and your industry.

1. Have someone in your organization serve on EIA and IEEE Committees. The best way is through membership, but EIA has an alternate arrangement on which I, or anyone else at EIA Headquarters, can give details.

2. When asked to comment on proposals, read them thoroughly and comment. When I see one returned "no comment" from someone who should be interested, it is difficult not to conclude "he did not read it", especially if it contains some obvious errors!

3. Get your comments in on time so they can be correlated or discussed.

4. If you see a need for a standard or want to propose one, notify the chairman of the appropriate group.

Standards, like death and taxes, are going to be with us, like it or not, so let us see that they are more pleasant than death and put to better use than taxes.

The previously mentioned documents can be purchased from the following sources:

IRE Standards - IEEE, The Institute of Electrical & Electronics Engineers, Inc.
345 East 47th Street, New York, N.Y. 10017

EIA Publications - Electronic Industries Association
2001 Eye Street, N.W., Washington, D.C. 20006

IEC Publications - American National Standards Institute
1430 Broadway, New York, N.Y. 10018

TIME CONTROL OF FREQUENCY SHIFT KEYED TRANSMISSIONS AT VLF

R. R. Stone, T. H. Gattis and R. C. Petit
Naval Research Laboratory

and

T. N. Lieberman
Naval Electronic Systems Command

Summary

Since 1960, the Navy has provided phase stabilized transmissions from its high powered VLF communications stations as a means for rating precision frequency oscillators at remote points. Recently, emphasis has been placed upon adding precision time reference capabilities to these transmissions.

A system has been developed by which the point of transition of the frequency shifted signal can be controlled at a precise rate and a defined time during normal communication periods. Means have also been developed to permit the transmission of scheduled time signals.

The time of the frequency shift transition, the phase coincident point between the MARK and SPACE frequencies and the zero crossing of the positive slope of the "ON frequency carrier are controlled to ± 1 microsecond of the station clock. Cesium beam standards and the associated clocks at the station are referenced to the U.S. Naval Observatory via "flying clocks" or the DSCS satellite time transfer system.

The first installation of this system has been made at station NWC at the Harold E. Holt Communications Station at North West Cape, Australia.

This paper discusses techniques, instrumentation and problems involved in the development of the system for the transmitters. Data on the operation of the system at the remote receive end will be presented and discussed at a later date when the system has been operated sufficiently long to permit evaluation.

* * * * * *

Since 1960 the Navy has employed its high-powered VLF system as a means of rating precision frequency oscillators at remote points. The wavelengths of these frequencies (15 KHz to 35 KHz) are sufficiently long, compared to variations in the length of the propagation path, that phase tracking of the received carrier at remote points, even after several reflections, can be easily accomplished. Atomic standards at the transmitter provide frequency control of better than one part in 10^{11} and permit the rating of oscillators at the received point to better than one part in 10^{10}.

At the present time, there are seven of these high-powered VLF transmitters as shown in Figure 1. (A recent installation has been made, NDT, in Yosami, Japan.) New antenna systems are being installed in Hawaii and Annapolis. Some of these stations have been operating since the mid-1930's, at which time they employed tuned circuits at the input and intermediate stages. At the present time, all stations have been updated with broadband amplifiers and they employ tuning only at the output/antenna. The newer system greatly simplifies the transmission of time signals.

Operation in the CW mode is quite simple, since all that is required is a precision reference for the carrier. Frequency shift keying (FSK) presents a somewhat more difficult problem. The format used for the VLF frequency shift signal is shown in Figure 2. The bandwidths of the antenna systems at these frequencies are narrow and they restrict the speed and magnitude of the carrier shifts. A 50-baud, 7.0 teletype code is employed. The bit lengths are 20 milliseconds (msec) and the transition time between the stabilized points of the carriers is 2 msec. Fifty cycle carrier shift is employed.

To permit the use of phase-coherent receivers for phase comparison at the remote sites, it is necessary that the carrier being measured be continuous in phase, as shown on the lower portion of Figure 2. Because of the high power involved and the high Q of the antenna system, phase discrepancies at the point of transition will produce transients which result in high voltage flash-overs in the transmitter. Where two carriers are employed, it is necessary that the transition between them occur at a point of phase coincidence. Fortunately, with bit lengths of 20 msec and carrier separation of 50 cycles, phase coincidence will occur at each transition point. However, at the time of the installation of the FSK system, it was not operationally feasible to precisely control the baud rate; therefore, one of the carriers was controlled in phase to maintain phase coincidence at the transition point and the other carrier was phase-controlled relative to the reference standard. In actual operation, the frequency of the phase-controlled carrier is set at the station assigned frequency and the other carrier is offset 50 cycles either above or below the assigned frequency. At the remote receive end, the on-frequency carrier is phase-stable except for propagation variations and the offset frequency carrier contains the phase variations which are required to compensate for the mis-timing of the teletype bit stream.

The instrumentation which is now available at all VLF stations is shown in Figure 3. A cesium beam reference standard is used to drive a divider bank which produces the frequencies needed in the synthesizer and also to provide two frequencies, 50 cycles apart, to be used by the keyer. These two frequencies are fed into a switch and also into a coincidence detector. The input key stream sets a gate which is activated by the coincidence pulses. The output of the gate in turn activates the switch in response to the input keying at the phase-coincidence points between the two frequencies. The shifted output from the switch is converted to a sine wave by a locked oscillator and controlled in shift time by the time constant of a servo-controlled oscillator. This output is mixed with the frequencies in the synthesizer to produce the VLF control frequencies of 14

to 36 KHz.

Recently, emphasis has been on controlling the point of transition of the FSK signal at a precise rate and a defined time. A system has been developed and was put into operation at the Northwest Cape, Australia (NWC) installation in January 1971. When a communications system is used for precision time and frequency purposes, it is necessary that the communications aspects of the system be preserved. In the case of the VLF system, the teletype code stream carrying the information is generated in a remote classified area and is "covered". Several sources of error must be recognized and compensated for by the system. First, noise bursts sometimes occur on the control line, producing extraneous bits or "hits"; second, at certain times erroneous or non-controlled (that is out-of-time) signals may be imposed on the line; and third, some mistiming may occur between the keying stream and the precision time stream generated from the standard.

A storage-retimer unit has been developed which overcomes these problems by storing the incoming teletype code stream, and reclocking the code stream out at a rate synchronized to the local frequency standard. The input and output code streams are identical except delayed in time.

A block diagram of the unit is shown in figure four. The input section consisting of the gate, count to twelve and flip flop eliminates noise bursts and transients from the incoming code stream. It is, in effect, a low pass filter which must have a count of twelve before it will acknowledge a change to mark or space. This circuitry quite effectively eliminates extraneous noise bursts. The input clock generator circuitry, consisting of the phase detector, two counts to fourteen, and a count to sixteen, recognizes an out-of-time or randomly keyed signal and generates a 50 Hz signal which is synchronized to the incoming code stream. The tracking range of the input clock circuitry is limited to a few parts in 10^{-3}. This prevents erroneous shifting of the incoming data and consequent loss of synchronization in the communication system. The incoming code stream is stored in the eight bit storage register and is clocked out by the 50 Hz standard clock pulses. It is expected that timing errors of several parts in 10^{7} will occur between the incoming bit stream and the standard clock controlled bit stream. There are approximately four million 20 milliseconds bits in a day, therefore one half of the storage or four bits will allow an offset of approximately 1×10^{-6}. The display section displays the storage position of the incoming data versus the outgoing data.

Since all of the frequency and timing aspects of the transmissions are derived from the same reference standard, it is possible to set the transmission on epoch time relative to the station clock. A phase shift servo control system, which is part of the frequency stabilization system, maintains the phase of the transmitted "on frequency" carrier in a fixed relation to the driving signal. The first step is to set the zero crossing of the positive slope of one of these cycles on epoch time, illustrated in figure 5. The center of the frequency shift transition was chosen as the time reference point. This point is moved by varying the phase of the off-set frequency until it is co-incident with epoch time. The center point was chosen because it can be easily set at the transmitter and extracted at the receiver by means of a discriminator. To date no conclusive long term data has been taken to determine the accuracy to which the time may be extracted at the receiver. However; the mid-transition point and the carrier cycle crossover are controlled at the transmitter at NWC to better than one microsecond. These points are set manually at present, but automatic equipment is being developed. The accuracy to which the mid-point transition can be extracted at the receiver is about $\pm$ 10 microseconds (μsec). This will allow the selection of a particular cycle since the period of the VLF frequencies are from 66 μsec. to 33 μsec. Extraction of time from the cycle zero crossing should be approximately what can be obtained by phase recording or $\pm$ 1 microsecond. Control of the phase of both carriers will also allow the use of a system somewhat similar to that used by the Bureau of Standards in which each of the carriers can be phase tracked separately, then mixed in a coincidence detector to yield 20 msec. markers. It can then be compared against the local clock.

The control of the transition of the FSK signal will provide time markers at 20 msec. intervals throughout the normal communication; however, for many cases, it is necessary to periodically identify seconds, minutes, and hours. The simplest method of accomplishing this is to periodically send time signals. Fig. 6 is a block diagram of the FSK time signal code keyer. It consists of a series of gate controlled outputs from a digital clock, such that "time" in the American code can be transmitted when desired. Twenty-five Hz. reversals are gated by the American time code to produce the keying signal for the FSK Coherent Generator. The beginning of the time signals is controlled manually and the shift back to normal FSK traffic is controlled by the one pulse per hour marker. The shift occurs one second after the hour. A diagram of the code pulse is shown in Fig. 7. The first three hundred msec. of the code consists of 20 msec. reversals. This is followed by a 700 msec. steady signal of the offset carrier. The beginning of the second occurs at the half transition point of the start of reversals. The time signal which is produced at the remote receiver is very easily recognized by ear. Figure seven shows the time signals to begin two minutes before 0430 and 1630. This is the format which is used at NWC. Starting of the signals at the half hour point was necessary because communications could not be interrupted at the hour point. It is expected however that at other VLF stations the time signals will be sent during the last five minutes before the hour.

Fig. 8 shows the format of the American time code. The 29th second is omitted from every minute, and there are seconds omissions according to this table which indicates the minute for the time mark. The time mark itself is followed by a 1 second tone, or in the case of FSK by one second of reversals.

Several techniques are possible for the detection of a frequency transition between two radio frequency (RF) carriers. Two which come to mind immediately are the frequency discriminator which is found in many communication type receivers and the synchronous detector which is used in VLF Phase Tracking receivers. Figure 9 illustrates the use of a synchronous detector for identifying a frequency transition point. The RF signal from the antenna is amplified and connected to one input to a synchronous detector. A 24 KHz signal synchronized to the local frequency standard supplies the second input. The detector output for a locally generated 24 KHz reversal signal is shown in the photographs on the left of Figure 9. Note that the negative (or positive) going cycle can be estimated to approximately $\pm$ 10 μsec. The two photographs on the right in Figure 9 show detected FSK signals received from NBA at 24 KHz. The negative going transition can be estimated to $\pm$ 10 μsec. The modulation on the detected waveforms is RF which was passed through the

low pass filter following the detector. It should be noted that the two lower photographs have been averaged over several seconds by the oscilloscope camera. Even so, it would not be difficult to visually obtain an accuracy of well within 100 μseconds directly.

Figure 10 illustrates the use of a frequency discriminator for detecting the transition. The RF from the antenna is amplified and mixed with the receiver local oscillator to obtain the immediate frequency (IF). The IF bandwidth must be sufficiently wide to pass the 50 Hz offset frequency. The IF signal is fed into a discriminator whose output is shown in the photographs in figure 10. The two photographs on the left are for reversals from a locally generated FSK (18.6 KHz) signal. The photograph in the lower left indicates the ability to detect a transition to approximately $\pm$ 10 μsec. The photographs on the right show the response for a FSK signal from NPG 18.6 KHz. The lower right photograph illustrates that the transition from NPG can be detected with approximately the same precision as for the locally generated reversal case. The two lower photographs were average over several seconds. The double pattern is caused by overlapping sweeps. The receiver input signal levels were approximately 25 μv. Locally generated signals in the 1 μvolt region have yielded approximately the same results as outlined above, but at present no off-the-air checks have been made for signals in this region.

In order to obtain the $\pm$ 10 μsec. accuracy needed to identify a cycle, some form of signal averaging is required. An experimental FSK - Time Signal Averager has been developed by NRL. A block diagram of the unit is shown in Figure 11. The unit measures the time delay between a local clock and a received transition, and performs an average of a number of time delay measurements and synchronizes the local clock pulse to the average.

The discriminator output is fed into the transition detector/shaper circuitry which produces a pulse for each transition.

The 50 Hz reset/start signal to the time delay counter circuitry causes the 1 MHz counter to reset to zero and start counting. The counter continues to accumulate a total until a transition pulse occurs which stops the counter, or until a 50 Hz reset/start pulse occurs which resets the counter and starts the measurement again. If a transition pulse stops the counter, the count is added in the digital integrator to the previous total. This count/add sequence continues until the number of transitions measured equals the number of samples desired. The following sequence is then initiated: (a) The accumulated total in the integrator is then loaded into the shift register divider (SRD) circuitry, (b) A divide signal from the Master control section causes the SRD to divide by the number of samples taken. (c) The computed average of the transition delay appears at one input to the comparator circuitry. (d) A signal "on time" with the reset/start pulse from the master control section causes the comparator counter to reset to zero and start to count. (e) A gate control signal from the master control section activates the comparator circuitry. When the count of the comparator counter equals the computed average from the SRD, a reset pulse from the comparator resets the master counter. The two master counter clock outputs (reset/start and the transition time estimate) are now synchronized to the average transition pulse (time marker). The transition time estimate and the reset/start pulse from the master counter have a 20 msec. repetition rate and are offset by exactly ten msec. The transition time estimate is synchronized to the averaged transition time. The number of samples can be programmed to average 2^4, 2^8, 2^{12} or 2^{16} transitions.

Figure 12 shows a sample of the results obtained using the receiver/discriminator output and the Time Signal Averager. This photograph was obtained using a locally generated 24 KHz reversal keying signal. The horizontal white line shows the spread of the incoming transitions at the input to the time delay counter. Most of the transitions occur in a 100 μsec. region. The transition time estimate (dots shown in photograph) averaged over 2^8 samples has a variation of approximately $\pm$ 10 μseconds.

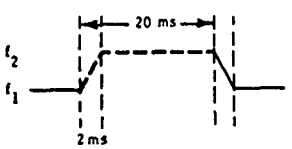
The repeatability of detecting the frequency transitions of an on-the-air FSK signal from NAA (17.8 KHz) is shown in Fig. 13. Using a communication receiver (discriminator technique discussed earlier) and the Time Signal-Averager, 2^8 transitions were averaged. Each 2^8 averaging period requires approximately 8 seconds with a psuedo-random keying signal. The data shown demonstrates the ability to measure the frequency transitions to within $\pm$ 10 microseconds. The signal level of NAA transmissions are high in the Washington area, but similar tests for transitions from NBA with signal levels received at NRL of approximately 30 microvolts have produced transition data in the same range as those shown for NAA. The deviation can be further reduced by increasing the averaging time. However, as presently constituted, without storage capabilities, the off-set frequency shifts too often to permit longer averaging times.

Within the next several months we expect to install time control equipment at NBA (24 KHz). On the basis of the results of data taken on these transmissions the various techniques will be evaluated. These results and associated techniques will be published later.

SCHEDULE OF TIME AND FREQUENCY TRANSMISSIONS ON VLF FROM U.S. NAVAL RADIO STATIONS

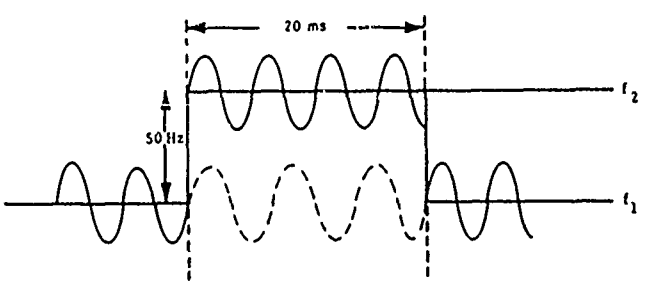
Station	Location	Frequency (kHz)	Nominal Radiated Power (kw)	Maintenance	Special Transmissions
NAA	Cutler, Maine 44°3819N, 67°1619W	17.80	1,000	1400 to 1800 UT each Friday	FSK for two hours followed by CW for one hour. Phase stable on 17.80 but not on 17.85 kHz.
NBA	Balboa, Canal Zone	24.00	150	1200 to 1800 UT each Monday	Time signals on CW Morse from 55 to 60th minute every even hour except 2355 to 2400 UT. FSK continuous at other times. Phase stable on 24.00 but no on 24.05 kHz.
NLK	Jim Creek, Washington 48°1211N, 121°5510W	18.60	250	1000 to 1500 UT second Thursday of each month	FSK continuous except five minutes before each even hour on locked key. Phase stable on 18.60 but not on 18.65 kHz.
NPM	Lualualei, Hawaii	23.40	1eu	1700 UT Monday to 0200 UT Tuesday 1st and 3rd Monday of each month.	FSK continuous. Phase stable on 23.40 but not on 23.45 kHz.*
NSS	Annapolis, Md.	21.40	85	1300 to 1900 UT each Wednesday	Time signals from 55 to 60th minute each hour. CW Morse continuous. Phase stable.
NWC	North West Cape, Australia 21°4910S, 114°0918E	22.30	1,000	0000 to 0300 UT each Monday	FSK and CW. Phase stable on 22.30 but not on 22.35 kHz.

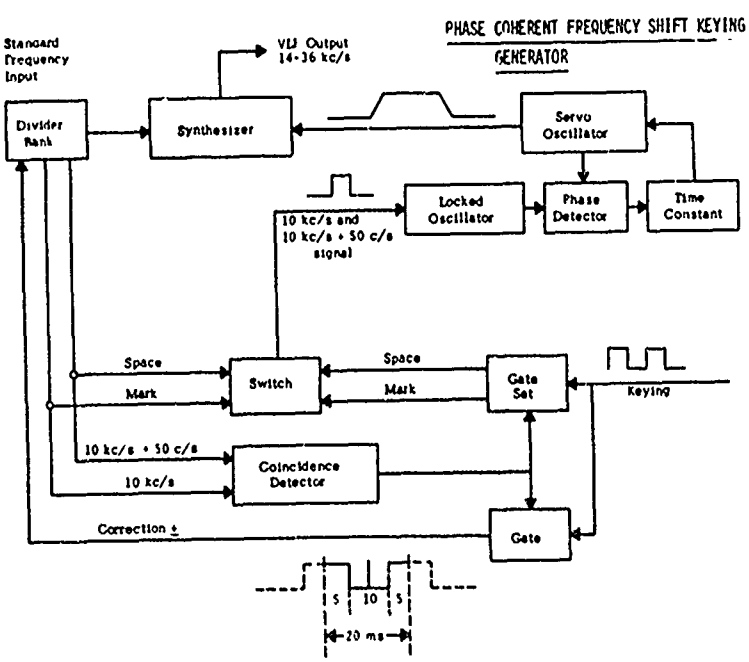
VLF FSK SIGNAL FORMAT



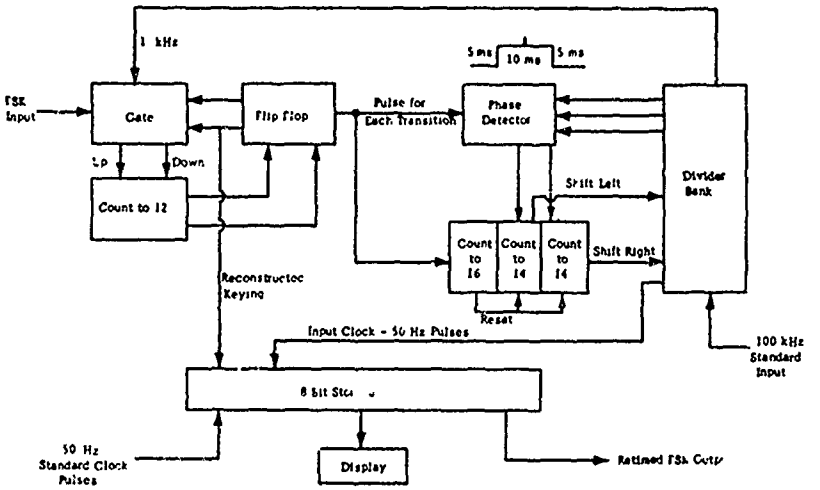
50 Baud
7.0 Code

Covered
50 Cycle Shift

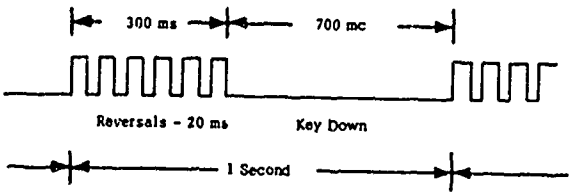




STORAGE/RETIMER

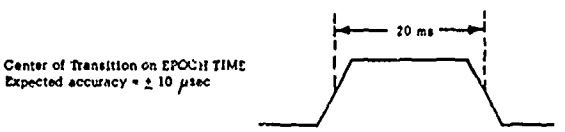


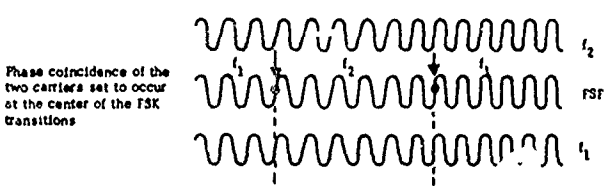
TIME CODE VLF



American Time Code - Begin 2 minutes before the even hour.

NEW TIME SIGNAL FORMAT





TIME CODE FOR U.S. NAVAL RADIO STATIONS (AMERICAN CODE)

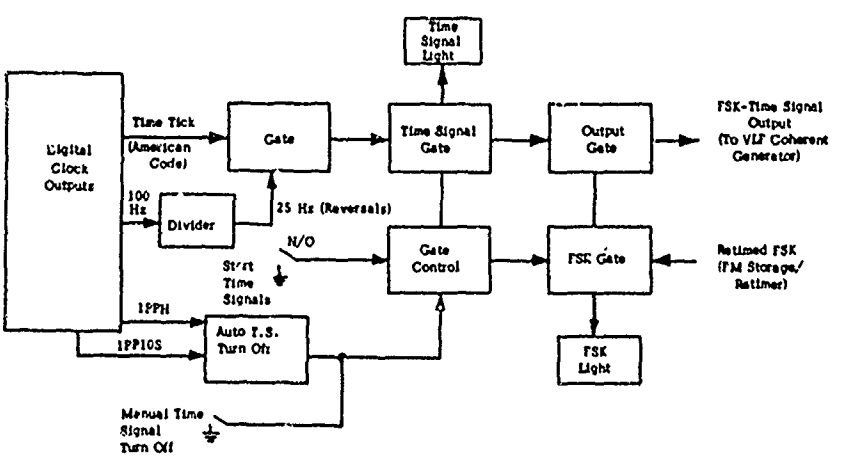
Time signals (dashes) are transmitted for each second of the five minute period with the following exceptions:

a. Omit 29th sec/min.

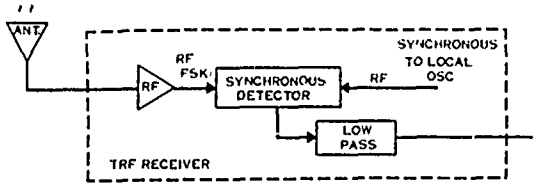
b. Second omissions permit minute identification.

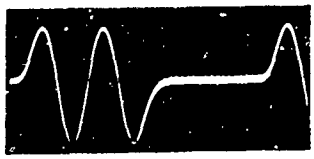
Minutes	Seconds										0 ⟶
	50	51	52	53	54	55	56	57	58	59	60
55	⎍		⎍	⎍	⎍	⎍					⎍
56	⎍	⎍		⎍	⎍	⎍					⎍
57	⎍	⎍	⎍		⎍	⎍					⎍
58	⎍	⎍	⎍	⎍		⎍					⎍
59	⎍										⎍

FSK/TIME SIGNAL CODE KEYER



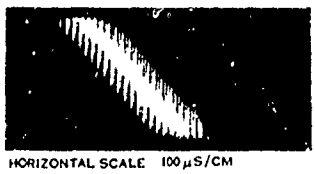
FREQUENCY TRANSITIONS USING A SYNCHRONOUS DETECTOR



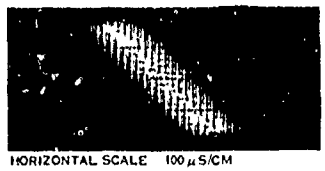


HORIZONTAL SCALE -5 MS/CM
SCALE EXPANDED ON NEG TRANSITION

HORIZONTAL SCALE 20 MS/CM
SCALE EXPANDED ON NEG TRANSITION

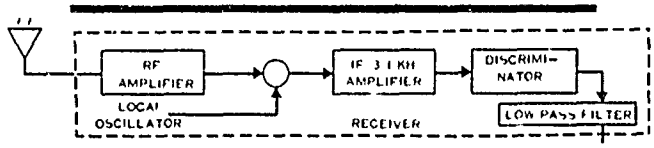


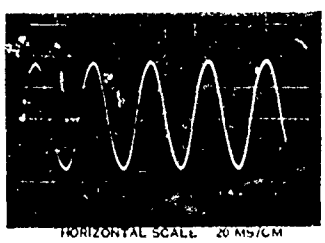
HORIZONTAL SCALE 100 μS/CM
LOCAL GENERATED
24 KHz REVERSAL SIGNAL



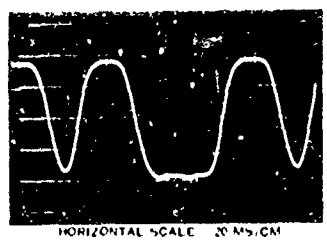
HORIZONTAL SCALE 100 μS/CM
FSK TRANSMISSIONS
FROM NBA 24 KH

FSK-TIME SIGNAL EXTRACTIONS USING A DISCRIMINATOR

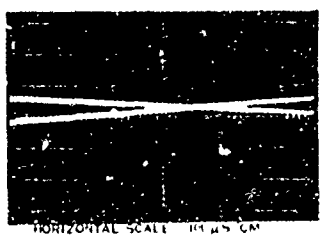




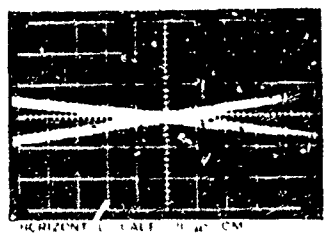
HORIZONTAL SCALE 20 MS/CM



HORIZONTAL SCALE 20 MS/CM

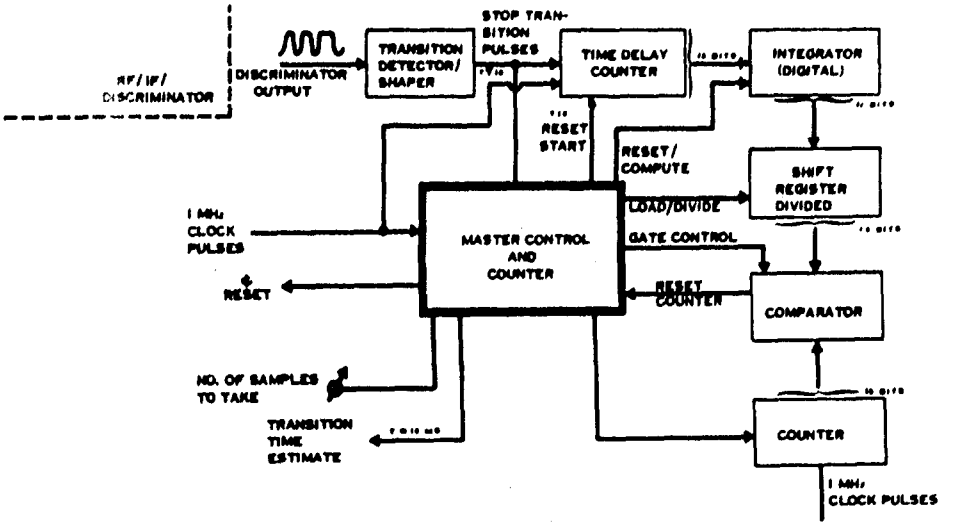


HORIZONTAL SCALE 100 μS/CM
REVERSALS FROM LOCAL FSK SIGNAL



HORIZONTAL SCALE 100 μS/CM
FSK TRANSMISSIONS FROM NBA NLF
24 KH AVERAGE OVER THREE ONDS

TIME SIGNAL AVERAGER



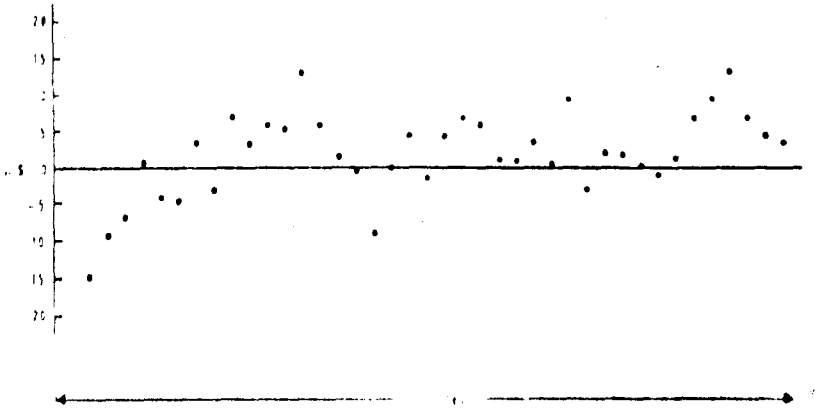
FSK-TIME SIGNALS AND 2^8 AVERAGES

HORIZONTAL: 40 μsec/cm

REVERSALS LOCALLY GENERATED (24 KHz)

REPEATABILITY EXPERIMENT
CUTLER, 17.8 KHz, 2^8 SAMPLES/ESTIMATE

15 APRIL 71



OMEGA VLF TIMING

E. R. Swanson and C. P. Kugel

Naval Electronics Laboratory Center
San Diego, California 92152

PREFACE

This paper is an abridgement of NELC Technical Report 1740[1] – a 143-page study done to consider application of Omega to timing of NASA telemetry stations.

SUMMARY

A timing site may be established by installing suitable equipment and selecting suitable signals for observation At the present time, a convenient choice of equipment would be a conventional navigation receiver with a whip antenna modified for signal-injection calibration using a special injector and keyer Path-selection studies can be made for the specific receiving site; however, daytime measurements near noon on the Omega transmissions between 10 2 and 13.6 kHz probably will prove most satisfactory Gross epoch can be established by lead-edge techniques. Epoch resolution then can be employed to identify particular rf carrier periods. Once initially set, epoch can be maintained by reliable dividers and periodic fine-epoch measurements at various carrier frequencies. If the signal stabilities are adequate epoch verification may be possible via signals from different stations at different frequencies. If significant differences are indicated, direct measurements for epoch verification should be initiated Clock frequency is derived from a combination of a quality frequency standard and a periodically adjusted phase shifter The frequency error of the standard is estimated by a regression of indicated epoch on time and is compensated by periodic phase shift. Adjustments are obtained from daily epoch errors derived over perhaps four propagation paths and are applied in accordance with an adopted adjustment procedure Through the use of a cesium frequency standard and a 60-day regression to deduce frequency, maintenance of change of epoch to better than 3 μsec and frequency to about one part in 10^{12} is possible. If special techniques are used, theoretical computations indicate the possibility of maintaining epoch continuously to better than 1 μsec.[1]

INTRODUCTION

DISCUSSION OF TIME

Time is one of the most tenuous concepts employed by man for precise measurements One unique feature distinguishing time from other units is the conceptual impossibility of building and maintaining a self-sufficient permanent standard Hence, while most standards can be manufactured and stored without subsequent intercomparisons of the quantities represented, clocks must be periodically intercompared to maintain precise time. Maintenance of precise time is as much a problem of techniques for time dissemination and adjustment as it is for a problem of construction of clocks.

The concept of time includes two major subdivisions - frequency and epoch. *Frequency* refers to the uniformity or periodicity of the time scale, *epoch* refers to the actual location or timing of periodic events. The concepts are developed more clearly in reference 2, where epoch is shown to be equivalent to the phase of a periodic function while frequency is the derivative of phase measured in cycles. Frequency or frequency stability always can be determined by periodic phase measurements or knowledge of phase stability and relevant autocorrelation processes for the medium. Appropriate relations are derived in reference 2.

Good time scales are based on regularly repeatable phenomena which are as periodic as practicable. However, *periodic* implies that each cycle of the event is indistinguishable from every other cycle. Hence, the epoch deduced from any periodic process is necessarily ambiguous by some unknown multiple of the period of the process. Any time determination thus also implies some method for *epoch ambiguity resolution* As timing becomes more precise, ambiguity resolution problems may become more severe, particularly if the clocks are separated. For example, the period of a 10-kHz vlf carrier transmission is 100 μsec. Hence, any time determination restricted to such a transmission would have an ambiguity of 100 n μsec where n is any integer. This ambiguity must be resolved by some other technique, such as comparison with a second periodic process of longer period – say, 300 μsec but which will differ in epoch from the original 10-kHz determination by less than 50 μsec. The ambiguity is thus resolved to an uncertainty of some integral multiple of the longer period, e.g., 300 m μsec The process may be repeated as often as necessary to resolve the coarsest time scale required, e.g., hours, days, or years. It is noteworthy that ambiguity resolution need not be performed with every epoch determination. Normally, some method is provided by which a count of the fine periods may be maintained, thus obviating the need for ambiguity resolution However, even if epoch is resolved properly initially, there is some inherent probability that [illegible]er counting will occur and hence a need for a method of ambiguity resolution Further, there is an inherent probability of error in resolving epoch ambiguity and hence a need for careful if occasional, attention to coarse epoch Reference 2 discusses epoch ambiguity resolution more thoroughly and stresses that ambiguity resolution errors are failures in the sense that a gross error will occur That is the accuracy of the system or technique is unaffected by ambiguity errors although the reliability may be critically dependent on ambiguity resolution

Dissemination of precise time thus implies a periodic process or processes and associated ability for measurement One such process is the phase of the electric vector of a cw radio signal If the transmissions are properly controlled and if frequencies and propagation paths are suitably chosen, properly interpreted phase measurements may be used for precise timing Although carrier phase measurements will be ambiguous by some integral multiple of the carrier period, the associated uncertainties may be resolved by other methods[3] or by additional transmissions at different frequencies[4]

OMEGA FORMAT DISCUSSION

Omega is a vlf navigation system scheduled for global implementation in the early 1970's The worldwide network will employ eight fixed ground stations to provide continuous redundant coverage anywhere in the world Each station will develop the radiated epoch from

a bank of four cesium frequency standards so that each station may be considered an equivalent source of precise time. Since hyperbolic navigation requires signals from a minimum of three stations, more information will be available than required for time dissemination to a known fixed location. Indeed, the four stations now operating provide global coverage for time dissemination although not for navigation.

The basic navigational frequency of Omega is 10.2 kHz. Transmissions of approximately 1-second duration are time-shared by each station within a 10-second commutation pattern. Navigation receivers are equipped with a commutator which, when properly set initially, can separate the transmissions from the various stations. Phase tracking of the 10.2-kHz transmission from each selected station is performed within the receiver. Typically, hyperbolic navigation is employed so that the phase tracks are paired and differenced in various ways to read out fractional cycle difference. Lane counters accumulate integral cycle crossings, or, alternatively, cycle count can be maintained by annotation on continuously operating strip-chart recorders. After the application of propagation corrections, two or more hyperbolic lines of position are plotted on specially prepared navigation charts. Intersection of the LOP's forms the fix. Normally, a fix may be obtained in under 3 minutes. Since the integral lane count is maintained rather than measured, a problem of lane identification may arise as a result of receiver malfunction, power failure, station outage, etc. Although lane resolution problems seldom occur and usually can be resolved by other means such as dead reckoning or independent fix, transmissions at two additional frequencies have been included so as to provide a method for lane identification. Omega is described in detail in reference 29.

Each station thus will transmit three frequencies on a time-shared basis every 10 seconds. The frequencies are 10.2, 11-1/3, and 13.6 kHz, any one of which may be used for navigation or timing. Of the available frequencies, 10.2 kHz has the longest period and greatest spacing between LOP's and hence presents the least problem from the viewpoint of possible navigational lane ambiguity or timing epoch ambiguity. 10.2 kHz is also the best calibrated of the Omega frequencies. However, 13.6 kHz will generally have the best signal-to-noise ratio and the greatest repeatability.

In addition to the three frequencies transmitted from each station primarily for navigation, two additional frequencies are being considered for intrasystem control transmissions of synchronization information. These two frequencies will be unique to each station and hence they may be received without commutation. As presently envisioned these control transmissions will be generated from the same timing equipment used to develop the navigational frequencies. They will range from 12 to 13 kHz and each station will be assigned two frequencies 250 Hz apart, e.g., Trinidad might transmit 12.0 and 12.250 kHz. Unlike the navigational frequencies, they will not be assigned specific segments but may vary within a commutation cycle for transmission of synchronization information. However, during each commutation cycle, each transmission will occur on at least one segment and the control frequency divisible by 100 Hz will occur on the segment immediately following the navigational transmissions. Possible formats are shown in reference 29, together with a more complete system description.

Omega transmissions will be almost ideal sources of precise time. Advantages include:

- Nearly continuous operation of each station
- Eight stations located conveniently around the world
- System synchronization procedures so that each station will be continuously held to agreed international time
- Extreme range from each station, providing redundancy
- Precise control of transmitted signals
- Stability of the propagation medium
- Transmissions of five related frequencies from each station
- Availability of propagation corrections

The primary disadvantage is that repeated measurements at various frequencies may be necessary to initially set or to periodically verify the coarse epoch. However, the necessary information is available from the transmitted carrier frequencies, the beat frequencies, and the commutation pattern.

INTERRELATION OF NAVIGATION AND TIMING

Using a system such as Omega, navigation and timing are inextricably interrelated. The navigation fixing problem using three stations may be viewed as the reduction of three navigation variates (signals) to deduce three unknowns - latitude, longitude, and time. In conventional navigation the time determination is disregarded. However, time is necessarily determined and, indeed, the time determination from the resultant fix position to any of the three stations must be identical. For example, the NELC synchronized Omega monitor at Cape Prince of Wales, Alaska, was being synchronized by a Naval Observatory 'flying clock' on 272100Z August 1969. The monitor position can be deduced from the actual Omega phase measurements of Norway, Hawaii, and New York as:

65° 37.2'N
168° 4 7'W

This indicates a position error of 0.6 nautical mile at a bearing of 34 degrees. The associated fractional cycle epoch error of the local clock is found to be a delay of 0.56 cycle. Simultaneous observatory measurement indicated a fractional cycle clock delay of 0.49 cycle, a discrepancy of 7 μsec. Note that although the velocity of light is 6 μsec per nautical mile, the fix error was 0.6 mile while the timing error was 7 μsec. It is often erroneously assumed that timing accuracy is directly proportional to fixing accuracy, actually the fix error can be large while the time error is zero.

Distinctions between Omega navigation and timing are thus not fundamental but only variations in application. For example, if time were precisely known, then a navigation fix could be used to obtain a more precise position estimate. Conversely, in the case of time dissemination to a fixed location, the position ordinarily would be known precisely and all available signals could be used to refine the time estimate. Although there is not a fundamental difference between navigation and timing applications, there are numerous practical differences. A receiver suitable for navigation is not necessarily suitable for timing nor, indeed, is a receiver suitable for timing necessarily suitable for navigation. Further, propagation corrections and accuracy estimates for navigation are not necessarily applicable for time dissemination or for estimating the accuracy of disseminated time. It makes no better sense to attempt to estimate practical timing accuracy from practical navigational accuracy than it does to attempt to estimate probable uncertainties in longitude knowing probable uncertainty in latitude but not the prevailing physics or geometry. However, valid estimates of timing capability can be made from proper application of knowledge of navigational signal predictability, stability, and correlation.

ACCURACY OF TIME DETERMINATION TECHNIQUES

PULSE TECHNIQUES

Pulse timing techniques have been used for many years. They have significant advantages over phase techniques employing multiple frequencies since there is no ambiguity problem except that associated with the repetition rate, which may be controlled by the system designer. Pulse techniques lack precision compared with phase comparison accuracies. Further, pulse techniques require relatively good signal-to-noise ratios.

Because Omega transmitting antennas systems have high Q and narrow bandwidth on the order of 10 Hz, the envelope of the transmitted wave form will approximate a rising exponential of the form $e(t) = E_{max}(1 - \exp kt)$. The exact shape will be different owing to other time constants in the transmitter.* A typical rise is shown in figure 1.

The rise is exceptionally slow compared to most pulse transmissions.

*Palmer uses a more general form stating that the rise is along an exponential curve 'more or less' of the form $e(t) = E_{max}(1 - kt)^N$ (reference 5).

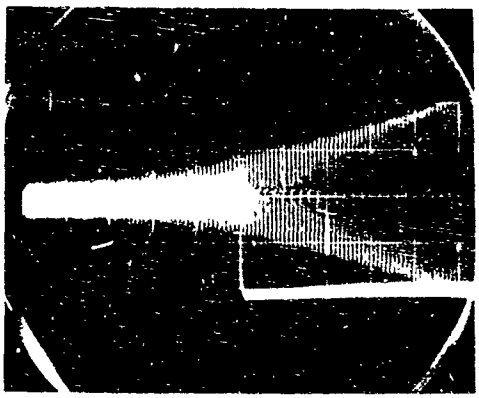


Figure 1. Haiku leading edge as received at NELC

TABLE 1. RESULTS OF LEAD-EDGE EXPERIMENTS.

Lead-Edge Date/Time Nov 1967	Intercept Cycles of 10 2 kHz With Respect to Reference Pulse	Uncertainty Coefficient q, Cycles	Signal Width, cm
031900Z	-12.87	2.57	0.43
062245	-15.70	0.65	0.12
unmarked	-16.03	0.80	0 12
032245	-16.75	0.47	0.20
022330	-16.94	0.58	0.21
131925	-17.97	1.43	0.22
211900	-15.59	0.72	0.20

A pulse inherently will be distorted by (dispersive effects) of the propagation medium through which it is transmitted. The resulting shape at a remote receiver is in general complex (a good discussion is given in Stratton, ref. 6). Usually, it is necessary to define the point on the lead edge at which measurements are to be performed and to specify how the measurements are to be interpreted before the propagation delay can be specified. Although little theoretical or experimental work has been done on the absolute predictability of the Omega lead edge, speculations are that the prediction problem will not be severe, owing to the limited frequency components present in the transmitted signal Further, the timing capability is not intended as a precise method and hence relatively crude predictions can be employed. If the portion of the lead edge actually measured can be considered to have an effective velocity between the group velocity and the velocity of light, then the total range in the prediction uncertainty would be only about 100 μsec over typical paths

No unusual instrumentation problems should be encountered in making absolute time measurements with the lead edge. The most severe problem is the poor signal-to-noise ratio, which can be overcome by photographic integration Delays may be expected in the antenna coupler and receiver but these can be determined from analysis of the circuits employed and should remain stable to the accuracy required Alternatively, a simple circuit can be constructed to simulate the Omega pulse and the receiving system can be calibrated directly

Vlf pulse timing has been investigated most recently by DePrins[7] who found a precision on the order of 100 μsec for a 500-km path to GBR and a 200-μsec precision on a 3000-km path to NBA Similar experiments at NELC have corroborated this increase in precision when the lower portion of the exponential rise is observable The experiment was restricted to measurements of the 10.2-kHz signal from Haiku as received at NELC The experimental arrangement included a receiver to obtain an unlimited rf within a relatively broad bandwidth of approximately 100 Hz This bandwidth was that of the available receiver and will have some effect on the observed rise times Eight-minute photographic integration was used on a normal oscilloscope display in which triggering was accomplished from a precision clock driven by a cesium frequency standard A reference pulse was superimposed High gain was used so that only the start of the 10 2-kHz pulse was observed It was then assumed that the voltage-time relationship was linear for the first 2 to 3 msec and a linear regression analysis was conducted to estimate the starting time (The analysis treated time as the dependent variable so that statistical expressions for the uncertainty of the ordinate intercept could be used) Seven lead-edges were analyzed The results are summarized in table 1

Two columns are of special interest The 'Intercept' column indicates the actual value of the time determination from each photograph as compared with a superimposed reference pulse from a General Radio clock (1 unit equals 98 μsec) The second column of special interest is the 'Uncertainty Coefficient q' Each timing estimate including its associated confidence limits is given by the intercept $\pm qt$ where t is the value of the t distribution for approximately 12 degrees of freedom and confidence limits desired Approximately 70 percent of the intercept should be within $\pm$q while 95 percent should be within $\pm$2q. Thus, the median value of q (approximately 71 μsec) should approximate the standard deviation of the intercept values (approximately 157 μsec). The disagreement can be explained by the 'width' of the trace used in the first determination The record apparently was overexposed so that the indicated value was shifted slightly Disregarding the first determination yields a median value of q of about 67 μsec while the standard deviation of the intercept is reduced to 90 μsec Although the agreement is sufficiently close considering the various uncertainties, it is interesting to speculate on the improvement which might be possible using better instrumentation The uncertainty coefficient reflects primarily variations on the photographs and scaling errors and linearization in the analysis 'q' is thus most sensitive to experimental technique while the standard deviation of the various intercepts must include not only experimental errors but also daily variations in propagation. If this interpretation is correct, then the primary errors in the determinations must be experimental rather than propagational.

One trailing edge was also investigated Trailing edge measurements nominally offer the advantage of better signal-to-noise ratio However, scaling difficulties were a substantial source of error and the very low slope possible on the scope display led to an intercept uncertainty of 130 μsec More elaborate instrumental techniques should lead to substantial improvement in accuracy

The Omega pulses thus offer a means of timing to a precision of about 100 μsec, using only crude instrumental techniques. It seems likely that improved instrumentation and analysis will yield higher precision

PHASE TECHNIQUES

The accuracy of epoch estimates determined from carrier phase measurements depends upon the accuracy of the predictions for the nominal long-term average phase expected as well as the temporal stability of the medium The errors may be described as a bias and a standard deviation which combine to yield a typical rms timing uncertainty However, the rms timing uncertainty is recognized to have two separate components The bias error will not contribute to errors in determining frequency and hence only the standard deviation is of importance Epoch errors, however, will directly reflect biases in prediction, epoch confidence cannot be improved beyond limitations imposed by prediction bias

Since modern Omega operation began in 1966, over 1 million hours of phase measurements directly applicable to estimating the timing accuracy of Omega have been obtained The data have been acquired primarily at monitoring sites associated with each transmitter The monitors continuously record the phase difference between signals received

from various remote Omega transmitters and the local transmitter. Since the propagation from the local transmitter to the associated monitoring site is predictable,[8] the local transmitter signal can be viewed as stable synchronized injection to the monitor receiver. Hence, Omega transmitting station monitors may be viewed as timing receivers. Additional data have been acquired from propagation experiments at Wales, Alaska, where a receiver using specially designed injection calibration from a cesium standard was operated for over 1 year.[28] The true epoch of the local reference was related to the Omega system epoch by repeated 'flying clock' calibration.

Similar data occasionally have been recorded at NELC in San Diego when the flying clock data have been available to relate the local epoch to that of the Omega system. Other direct comparisons are available as a result of flying clocks between the U.S. Naval Observatory and various Omega stations.[9]

Figures 2A and 2B exemplify a typical comparison between a long-path phase measurement and prediction. Observed phase differences normally are made hourly with time constants appropriate for navigation. The data are flagged to indicate measurements possibly affected by

OBSERVED PHASE DIFFERENCE (CENTICYCLES) 10.2 KHZ WALES, ALASKA *C- MAY 69

DATE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	N	AV	SD
									N	N	N	N	N								D	D	D	D			
8 MAY 69																		67	63	60	57	56	55	55	0	0.0	0.0
9 MAY 69	55	57	61	68	75	88	96	02	10	11	13	13	12								56	55	52	53	5	11.8	1.2
10 MAY 69	54	57	61	67	74	89	01	04	09	09	10	15	11	08	94	68	67	63	60	58	55	54	54	53	5	10.8	2.2
11 MAY 69	55	58	62	65	72	86	00	10	09	09	17	17	14	11	98	66	68	65	60	57	55	54	54	54	5	13.2	3.6
12 MAY 69	55	54	59	64	70	86	96	07	10	12	13	11	09	03	87	65	66	63	60	57		54	53	52	5	11.0	1.4
13 MAY 69	53	55	57	61	69	78	92	97	05	09	10	10	08	98	87	57	P60	P57	P55	P55	P54	P53	P52	P52	5	8.6	1.5
14 MAY 69	P53	P54	P57	P61	P68	P80	P90	P96	P99	P04	P11	P11	P10	P04	P90	P68	P65	P62	P59	P56	P53	P52	P52	P51	0	0.0	0.0
15 MAY 69	P51	P55	P59	P63	P71	P86	P92	P92	P04	07	06	02	02	97	88	61	61	56	53	52	49	49	46	51	4	4.2	2.3
16 MAY 69	52	54	57	61	S68	S82	88	97	02	03	05	06	05						58	52	53	50	51	53	5	4.2	1.5
17 MAY 69	52	54	56	59	66	82	91	97	05	10	14	12	11	06	85	63	63	59	55	S43	S51	52	52	52	5	10.4	3.0
18 MAY 69	53	46	S45	55	64	79	88	94	03	07	11	08	04	94	S82	S62	S63	S53	S56	51	S51	S50	53	53	5	6.6	2.9
19 MAY 69	52	55	58	63	67	77	91	96	01	04	08	07	06	01	S82	60	59	58	55	51	47	47	50	S51	5	5.2	2.5
20 MAY 69	S45	50	51	53	59	73	84	94	01	03	03	03	03	98	82	66	S65	60	59	S48	53	53	S49	50	5	2.6	0.8
21 MAY 69	50	48	56	56	63	77	86	90	00	05	07	06	01	94	82	61	S62	59	55	52	49	47	46	47	5	3.8	2.8
22 MAY 69	48	47	53	58	57	72	87	94	01	09	11	10	07	98	84	58	62	60	S58	S39	49	50	50	51	5	7.6	3.6
23 MAY 69	53	54	57	61	66	76	92	01	06	13	15	13	09	99	79	63	63	59	56	54	53	53	52	51	5	11.2	3.2
24 MAY 69	53	56	59	63	69	S81	89	97	03	07	S07	08	08	01	84	65	66	62	58	55	53	53	52	54	5	6.5	1.9
25 MAY 69	52	52	55	58	63	76	87	93	01	03	09	12	09	03	84	62	S61	S57	54	52	S41	47	47	49	5	6.8	4.1
26 MAY 69	50	49	54	58	64	74	87	91	01	06	11	11	09	00	83	64	66	63	57	56	54	53	51	52	5	7.6	3.8
27 MAY 69	S44	53	57	62	67	77	89	98	05	07	08	15	14	04	88	62	64	62	60	58	56	55	54	55	5	9.8	4.0
28 MAY 69	56	58	61	64	69	81	96	04	08	09	09	09	S09	99	S83	65	63	61	59	58	54	48	53	51	5	8.7	0.4
29 MAY 69	47	55	58	63	61	77	89	02	09	12	15	11	09	03	S86	68	68	63	62	S36	52	51	53	53	5	11.2	2.2
30 MAY 69	51	53	56	60	67	77	85	98	09	12	13	12	07	99	84	64	64	60	59	55	54	52	53	52	5	10.6	2.2
31 MAY 69	54	56	58	62	70	78	93	01	06	09	11	14	13	05	86	64	66	62	59	S39	52	S34	46	53	5	10.6	2.9

GMT

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
								N	N	N	N	N								D	D	D	D
19	21	20	21	20	19	21	21	21	22	21	22	21	20	16	19	15	18	19	16	18	20	21	21
52.4	53.4	57.3	61.0	66.6	79.1	90.8	98.4	5.0	8.0	10.4	10.2	8.1	1.0	85.9	83.3	64.4	61.2	58.0	54.9	52.8	51.6	51.3	52.1
3.2	3.4	3.7	3.7	4.5	4.6	4.6	5.1	3.4	3.0	3.4	3.7	3.5	4.3	4.2	2.9	2.4	3.1	2.6	6.7	3.6	4.5	2.7	1.8

SMC

3	0	-2	-5	-8	-24	-45	-54	-54	-54	-54	-54	-54	-54	-48	-10	-7	-4	-1	1	3	4	4	3
3	1	-2	-4	-7	-18	-39	-51	-54	-54	-54	-54	-54	-47	-23	-8	-5	-2	0	2	4	5	5	4

RMS

5.2	5.6	5.4	4.0	4.3	3.6	8.9	12.7	8.4	5.6	4.2	4.5	5.7	7.2	7.7	4.8	2.2	2.8	2.6	8.4	4.4	5.4	3.7	3.3

Figure 2A

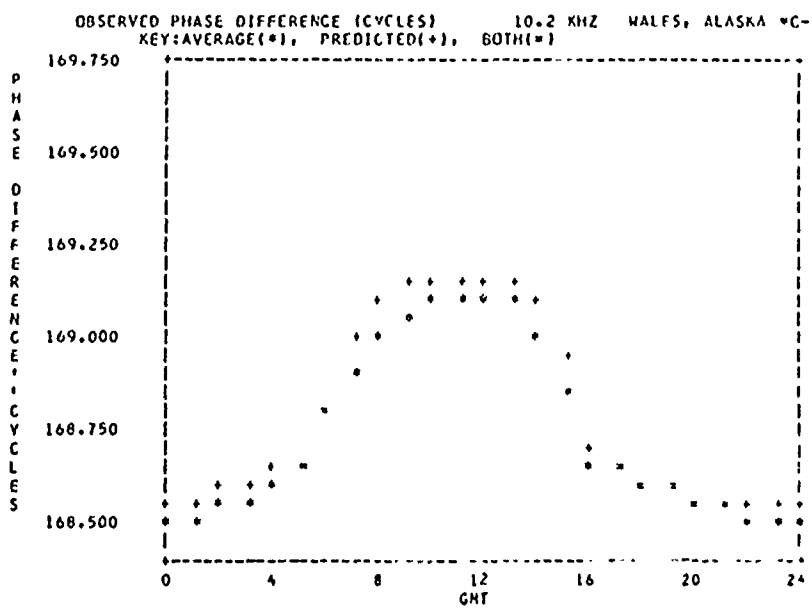


Figure 2B

SID's (S) and PCA's (P). The last eight rows (under GMT, SWC, and RMS) are, respectively, Greenwich mean time (1 = 0100Z, etc.); flags indicating day or night periods (i.e., when the propagation path is completely sunlit or dark); number of observations (less SID's and PCA's); averages (less SID's and PCA's); standard deviation (including SID's but excluding PCA's); two semimonthly sky wave corrections; and the rms value of the observed phase error with respect to prediction. Statistics are summarized in table 2, which contains a representative sampling of the quality of data to be expected, and indicates a sample standard deviation of about 3.8 μsec over 24 hours.

Short paths also may be used for timing under certain conditions. Prediction is extraordinarily difficult at short ranges where various propagation modes may be interfering. At certain distances, however, the resultant phase may be stable. If the nominal phase value can be determined either by direct calibration or by deduction using other propagation paths, the path can then be used for epoch measurement. An example is the observation of Omega New York at the U.S. Naval Observatory. Although the recorded value is about 40 μsec from anticipated, the measurement is stable.[10, 11]

FREQUENCY ESTIMATION TECHNIQUES

As described previously, frequency estimates are best obtained by making a regression of phase on time using periodic measurements of estimated epoch error. Usually, long-term estimates are desired, in which case the time interval over which the regression is conducted is long with respect to the inherent autocorrelation period of the medium. In any case, the frequency is obtained from

$$f = \frac{12\sum_{i}^{n} i\phi_i - 6(n+1)\sum_{i}^{n}\phi_i}{Tn(n^2-1)} \qquad (1)$$

where ϕ_i = phase estimate at the i^{th} interval in cycles

T = period between measurements in seconds

n = number of phase measurements

as previously given and derived in reference 2. If the epoch estimates are made sufficiently far apart that they are uncorrelated, then the accuracy of the frequency estimate obtained from equation (1) is derived in reference 2 as

$$\frac{\sigma_f}{f} = \frac{\sqrt{12}}{T\sqrt{n(n^2-1)}}\sigma_t \qquad (2)$$

when σ_t is the standard deviation of epoch measurements in seconds

TABLE 2. SUMMARY OF MEASUREMENTS.

f	Monitor Site	Time Period	XMTR	LOP	Night			Transition			Day			24 Hours		
					N*	RMS	σ	N	RMS	σ	N	RMS	σ	N	RMS	σ
10.2 kHz	Rome, New York	Jun 69	Trinidad	129 165	268	4.34	2 38	149	5.60	2.42	285	3.63	2 55	702	4.37	2.46
	Wales, Alaska	Aug 69	Norway	181.696				502	7.66	3 77	130	7.56	2.73	632	7 64	3.58
		May 69	Hawaii	168 586	109	5.81	3.45	309	6 42	4.10	87	4.28	3.32	505	5 99	3.84
		Oct 69			341	3 35	3 17	154	15 06	10 54	241	7.37	6.07	736	8.39	6.32
								10 2 kHz Median	cec		5.83	3.03		6.81	3 71
									μsec		5.	3.0		6.7	3.6
13.6 kHz	NELC, San Diego, California	Dec 69	Trinidad	279.089	80	6.03	3.98	101	8.99	4 08	24	7.41	1.21	205	7.80	3.82
	Pyramid Rock, Hawaii	May 69	Norway	362.385	-		-	574	8 75	4.43	83	8.48	4 76	657	8.72	4.47
	Rome, New York	Oct 69	Hawaii	351.895	217	9 01	5.47	365	7 62	7 46	148	5.18	5.07	730	7 64	6 48
	Wales, Alaska	Oct 69	New York	264 234	246	9.36	7.84	301	9 71	10.65	182	9 64	9.19	729	9 58	9 42
								13 6 kHz Median	cec		7.95	4 92		8 26	5.46
									μsec		5.8	3 .		6 1	4 0
								Average	μsec		5 8	3		6 4	3 8

*Number of samples

If the phase measurements are accurate to 1 μsec, an accuracy of one part in 10^{11} can be obtained after 2 days while a precision of 10^{-12} is obtained after about 10 days and 10^{-13} is obtained in 2 months. A plot of the anticipated precision in frequency as a function of the observation period is given in reference 2 (fig. 2).

Equation (2) applies only when the variations in the epoch estimates are not autocorrelated. For example, seasonal changes in propagation at 10.2 kHz, if not corrected with appropriate propagation predictions, can yield an apparent frequency error of 2×10^{-12}. Obviously any seasonal phase changes not removed properly by predictions will be correlated over long periods and will introduce frequency error. Similarly if the observations times were chosen improperly and significant diurnal changes were occurring, large frequency errors could be introduced.

Phase stability quoted in the previous section, 3.8 μsec, may be used with equation (2) to calculate the accuracy of frequency estimates derived from a single daily Omega epoch measurement where the observation time has been selected at random. Table 3 gives results for relevant frequency stabilities.

TABLE 3. ACCURACY OF FREQUENCY ESTIMATED FROM REPEATED OMEGA EPOCH MEASUREMENTS (RANDOM OBSERVATION TIME).

Frequency Stability	Observation Duration, days spanned
10^{-11}	6
10^{-12}	26
10^{-13}	133

Were the observation time favorably chosen, techniques employed to eliminate anomalous data or to introduce redundancy would improve the anticipated frequency stability. Note that fixed prediction or instrumentation errors do not introduce errors in frequency.

Direct comparisons between frequency measurements by flying clock and frequency estimates obtained from long all paths have also been made by Kugel using the specialized case of the Omega interstation paths.[9] Since measurements are made at both ends of these 'reciprocal' paths, calibration of reciprocal propagation phenomena is possible. Thus, the epoch errors are of higher quality than normally would be expected. Discrepancies between frequency estimates using a 60-day regression and short-term (2-3 day) measurements by flying clock were less than one part in 10^{12}. The discrepancies are believed due primarily to uncertainties in the short-term measurements and experimental complications rather than the radio determination.

Although of little interest at a timing site in continuous operation, techniques also are available to make short-term frequency estimates such as might be needed in an emergency resulting from gross malfunctions. In this case, measurements are made over a duration short with respect to the autocorrelation period of the propagation medium. As shown in reference 2, we might expect to be able to obtain frequency estimates accurate to about one part in 10^{9} in 15 to 20 minutes using a propagation path stable to 1 μsec. Wright made direct measurements using Omega signals received in Austin, Texas, to deduce the accuracy of short-term velocity estimates.[12] Although Wright's interest was in deducing the velocity and he made no mention of the accuracy in frequency estimation, the two are equivalent, since

$$V = \frac{(\Delta\phi)\lambda}{t} \qquad (3)$$

where $\Delta\phi$ is phase change in cycles

λ is wavelength

and t is time separation between observations

Since Wright obtained 0.2 knot during the day and 0.5 knot at night for a 15-minute observation span, the observed phase changes must have been 0.35 and 0.9 centicycle, i.e., 0.3 and 0.7 μsec, respectively. The corresponding frequency estimates are thus accurate to about 5×10^{-10}. Direct measurements thus show a better capability for short-term frequency measurement than indicated in reference 2.

AMBIGUITY RESOLUTION PROBABILITIES

Ambiguity or epoch resolution does not have intrinsic accuracy characteristics but instead must be considered on a probability basis. The relevant quantity is the probability that a certain epoch (3.4, 1.133 kHz etc.) can be correctly identified within another larger but possibly ambiguous epoch interval. For example, a user with only the three simple carrier signals would need some additional information in order to resolve the ambiguity of the 1.133 epoch. Because such information from other Omega frequencies, pulse techniques, satellite updates, etc., most likely will be available, this section will concentrate on the resolution probability for the 3.4 kHz.

Such resolution, of course, will depend directly on the ability to predict the carrier phase delays at the various monitor sites. Assuming that prediction capability is constantly improving and that results from past operation will be applicable to the future, we may consider some statistical studies undertaken to investigate lane resolution probabilities. These studies are summarized in references 13, 14, and 15, which present results based on thousands of hours of phase data taken at Omega frequencies at several sites around the globe. Although the reader may consult the references to assess the significance of the statistics to the data under consideration, the overall impression is that lane resolution is possible under all but a few types of serious propagation conditions, notably, PCA and SID onsets and large diurnal transitions. The ability to maintain the difference frequency epoch under disturbed conditions depends not only upon the behavior of the ionosphere, but also upon the ability of different receiver channels to respond similarly and of recorded tracks to indicate the same time on all frequencies. Some of the lane resolution failures found in the aforementioned studies possibly resulted from data being taken on different receivers and recorders, which may cause serious problems during rapid phase changes or high noise conditions. Experience has shown that Omega receivers exhibit a wide range of dynamic response depending upon integration time constants and sensitivity.[16,17] Such variations may produce significantly different behavior during large PCA onsets which have been known to exceed 1 cycle of carrier phase in a few minutes. The user likely will be aware of disturbances as they occur and will not worry about apparent lane loss at those times. However, without some additional warning service to corroborate the existence of a disturbance, the sight of all carrier phase tracks moving downscale simultaneously might lead the user to suspect his equipment, and some additional checks could be made. If the monitor is on the sunlit side of the earth, all paths are partially sunlit and would react to a disturbance. The difference in the reactions distinguishes clock failure from an SID, as some paths will be affected more than others.

If possible epoch loss during disturbances can be avoided, or at least recognized, there still remains the problem of prediction errors during undisturbed times. Evidently, if the 10.2 kHz carrier phase cannot be predicted to within 50 cec, the correct 10.2 lane cannot be established. However, such data would not be of use for fine epoch estimation but could still be used to resolve the 3.4 epoch provided the 13.6-kHz phase has a similarly gross prediction error (also of the same sign). Long path measurements between the Omega stations indicate that the direction, if not the magnitude, of prediction error bias tends to be similar on all frequencies. Possible exceptions are polar paths, which only recently have received the attention they warrant,[18,19,28] and equatorial propagation in general.[20,21]

Note that there are two separate and distinct needs for epoch resolution at a timing site - initialization and verification. When first established, the timing site must determine crude epoch. For example, the initial epoch determination of the 3.4-kHz (294 μsec) period from 10.2- and 13.6-kHz measurements must rely on propagation predictions. However, the initial measurements can be made over several propagation paths during stable periods so as to minimize the effects of prediction errors. Since the probability of an epoch error in a daytime measurement over a single path is less than 10^{-3}, the probability of obtaining the same epoch error over several paths under stable conditions should be vanishingly small. Once the epoch has been set initially, continuous measurements can be made and any residual prediction differences noted. Therefore, subsequent epoch verifications need not be degraded by differential prediction errors between the various frequencies. In essence a 24-hour capability for epoch verification or reestablishment is obtained.

The probability of an epoch resolution error can be reduced almost to an arbitrarily small value. The technique of 'selective resolution' demands that a 'good' epoch match be obtained from the two measurements being employed to resolve epoch. For example, if a 10.2-kHz carrier phase of 0 cec is indicated, we should like the 3.4-kHz beat to measure 0, 33, or 67 cec exactly. If the actual measurement is 16 cec, then the best estimate would be that the epoch was in the first period of the 10.2 kHz, but the estimate would be nearly a guess, since a difference measurement of 17 cec would have yielded the second 10.2-kHz period as the best estimate. By demanding close epoch agreement, we reduce the probability of epoch resolution, since no choice will be forced under marginal conditions. However, the probability of *incorrect* epoch resolution is vastly reduced. During the day, it is possible to reduce the probability of correct 10.2-to-13.6-kHz resolution from 0.999 to 0.998 while reducing the probability of incorrect identification from 10^{-3} to 10^{-5}. If even closer agreement is required, the probability of error can be further reduced.

Since the epoch resolution capability is due primarily to the high correlation of phase fluctuations at different frequencies over the same propagation paths, attempting an epoch resolution using measurements on different frequencies over different paths will generally result in an excessive number of failures. However, if the phase measurements normally made to adjust fine epoch are taken during favorable periods and if residual prediction errors are removed, the accuracy and correlation of the measurements over two separate paths might be adequate for continuing epoch verification. If the coarse epoch is indicated as questionable, then epoch resolution measurements over common paths can be made. Otherwise, the epoch would be verified by the usual adjustment measurements. This procedure reduces the number of measurements required while still yielding a significant number of relatively independent observations. Of course, personnel making the calculations must recognize the possible implications of large disagreements in epoch estimate.

ADJUSTMENT PROCEDURE

Maintenance of proper time requires procedures by which the epoch of the local clock can be adjusted to synchronism with external reference information. As previously described, the receiving equipment from which the epoch estimate is derived will include an antenna coupler (if required), phase tracking receiver, and suitable calibration injection or reference generation equipment. Inputs to the receiving complex thus include a time reference from the local clock and the radio signals. The output will be phase comparisons which are processed to obtain estimates of the error in the local epoch. The clock is independent of the receiving complex and may be illustrated schematically as in figure 3. An adjustment procedure is thus a sequence of operations from which periodic changes in the phase shifter position are derived such that the clock output is nearly correct as often as practicable.

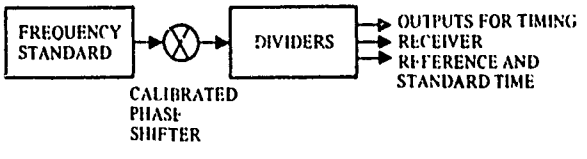


Figure 3 Schematic diagram of clock

Timing synchronization has been and still is being studied extensively in its application to the Omega system,[23-27] which represents an attempt to maintain several remote clocks in agreement with each other and also with an external reference in the most convenient and optimum manner. The formulations presented here are some of the results of these studies which are applicable to the general remote clock synchronization problem. This section contains only a brief outline of the suggested procedures while reference 1 contains the complete algebraic formulations. The procedure is a specific one which presently is being used to synchronize the Omega system and includes four hourly adjustments for frequency offsets and weekly phase corrections for indicated epoch errors. The frequency estimates are based on a 60-day least squares regression of nine data points 1 week apart. The procedure is convertible to other regression schedules and waiting periods between epoch adjustments.

The starting point for this discussion is an equation relating the epoch* of a clock at the present time t to some past time t-1

$$\phi_t = \phi_{t-1} + E_t + F_t T \tag{4}$$

where ϕ_t = phase epoch at the end of the t^{th} interval

ϕ_{t-1} = phase epoch at the end of the $(t-1)^{th}$ interval

E_t = phase correction applied during t^{th} interval

F_t = frequency radiated over t^{th} interval

T = interval length

Equation (4) states that the epoch that existed at time t-1 has been altered by a discrete adjustment inserted by the timekeeper and also by the accumulated phase offset produced by the frequency difference between the local clock-plus phase shifter combination and the 'perfect' reference. This frequency difference is given by

$$F_t = f_t + \delta f_t \tag{5}$$

where f_t = cesium standard frequency offset from perfect reference over t^{th} interval

δf_t = negative of the estimated cesium frequency offset, or 'accumulation,' over t^{th} interval

δf_t may be estimated by a least squares regression of phase on time.

If no adjustment was being made for the inherent frequency offset of the cesium standard, the term f_t would cause the epoch to advance or retard depending on whether this offset was positive or negative. Because the estimate term δf_t is computed from phase shifter adjustments intended to remove the apparent cesium offset, the regression line will have a slope opposite in sign to the offset and the addition of the two terms in (5) should cancel one another and result in no change to the epoch. In reality, the estimates are imperfect and will result in some indicated epoch discrepancy at the end of every interval. The correction term E_t is designed to remove this error and is given by

$$E_t = \theta \hat{\phi}_{t-1} = \theta(\phi_{t-1} + N_{t-1}) \tag{6}$$

where $\hat{\phi}_{t-1}$ = the estimated epoch at the end of the $(t-1)^{th}$ interval

ϕ_{t-1} = the true phase epoch at the end of the $(t-1)^{th}$ interval

N_{t-1} = the noise on the estimate $\hat{\phi}_{t-1}$

θ = parameter to be found

The parameter θ is considered to be a 'believability,' or weighting factor which permits the flexibility of using the known stability of the standards as a guide for interpreting the real epoch error as something different from 100 percent of the indicated estimate $\hat{\phi}$. Because minimization of the variance of the epoch errors is desired, a result of θ as a negative fraction is expected from this exercise. The procedure to be followed is recursion and minimization of variance for equation (6), with θ as the controlling parameter. The results shown below employ a frequency recursion of a 60-day least squares estimation.

The plots shown in figure 4 were obtained from a graphical analysis of several var ϕ_t versus θ plots as shown in reference 1. $\theta_{optimum}$ is the apparent minimum of those curves while $\sigma_{\phi_{min}}$ is related to the value of var ϕ_t resulting at those minimums. The figure indicates that

*The epoch terms used here may refer to phase or time as long as all quantities have the same units in the equations.

$\theta_{optimum}$ varies inversely with noise and directly with oscillator stability, as expected. The plot of $\sigma_{\phi_{min}}/\sigma_f$ is intended to indicate the linear variation of epoch with noise for any standard variability, when the optimum θ is being applied; e.g., when the noise is five times the standard variability, the application of a θ of -0.3 results in an epoch variability of only about 3.5 times worse than the standard. The apparent extrapolation to 0.1 indicates that under conditions of zero noise, the epoch variability can be no smaller than that of the standard.

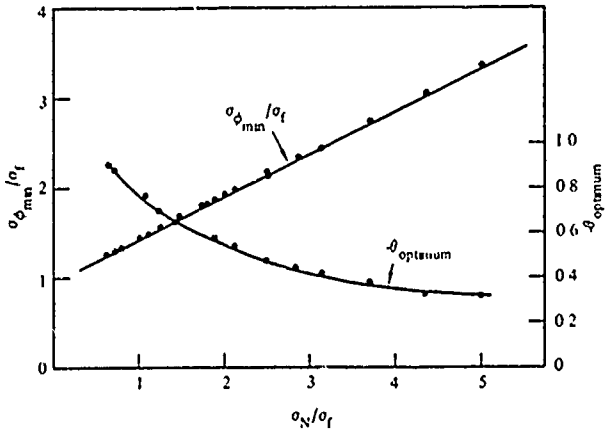


Figure 4 Control procedure analysis results

CONCLUSIONS AND RECOMMENDATIONS

By means of the techniques described, epoch at timing sites may be deduced and maintained to better than 3 μsec and frequency can be maintained to about one part in 10^{12} via Omega transmissions. Theoretically, employing special techniques should make possible the maintenance of epoch to better than 1 μsec. Prior to implementation of a complete global Omega system in the early 1970's, a useful timing capability presently is available from the existing four stations although at reduced accuracy in remote areas. The procedures described are essentially those which have been employed in synchronizing Omega for the past 4 years, and hence may be recommended as practicable for immediate implementation.

However, much additional work remains to be done. The areas requiring special attention include pulse timing, receiver installation design, propagation prediction, interpath cross correlation and signal autocorrelation, adjustment procedure optimization, and autocorrelation of the frequency variations of quality standards. The correlation questions are potentially the most important and most difficult. Conflicting data are presently available and even relatively low long-term correlation in frequency standards can drastically change the optimization of the adjustment procedure. Similarly, cross correlation of phase fluctuations on various paths may be significantly different for timing applications than for navigation owing to the elimination of recognized disturbances.

REFERENCES

1. Naval Electronics Laboratory Center TR 1740, OMEGA VLF Timing, by E. R. Swanson and C. P. Kugel, 5 November 1970

2. Swanson, E. R., "Time Dissemination Effects Caused by Instabilities of the Medium," p. 181-198 in North Atlantic Treaty Organization, Advisory Group For Aerospace Research and Development, Electromagnetic Wave Propagation Committee, Phase and Frequency Instabilities in Electromagnetic Wave Propagation, Editor K. Davies, Slough, England, Technivision Services, 1970 (AGARD Conference Proceedings No 33)

3. Shapiro, L. D., "Time Synchronization From LORAN-C," IEEE Spectrum (Institute of Electrical and Electronics Engineers), v. 5, p. 46-55, August 1968

4. Chi, A. R. and Witt, S. N., "Time Synchronization of Remote Clocks Using Dual VLF Transmissions," p. 588-612 in U.S. Army Electronics Command, Proceedings of the 20th Annual Symposium on Frequency Control, 21 April 1966

5. Palmer, W., Contract N00024-67C-1416; OMEGA Technical Note 17, Proposal For Standard Time and Frequency by OMEGA, December 1969

6. Stratton, J. A., Electromagnetic Theory, Section 5-18, McGraw-Hill, 1941

7. Guisset, J. L. and others, "Réception de Signaux Horaires sur Ondes Myriamétriques," Bulletin de la Classe Des Sciences (Academie Royale de Belgique), 5th Series, v. 52, p. 490-499, 5 March 1966

8. Watt, A. D., VLF Radio Engineering, Appendix C, Pergamon Press, 1967

9. Naval Electronics Laboratory Center Report 1529, OMEGA System Synchronization, by C. P. Kugel, 10 January 1968

10. U.S. Naval Observatory, Daily Relative Phase Values (Issued weekly)

11. U.S. Naval Observatory Time Service Letter, Instructions to Interpret the 'Phase Values' Bulletins and Teletype Messages, 30 September 1968

12. Wright, J., "Accuracy of Omega/VLF Range-Rate Measurements," Navigation, v. 16, p. 71-79, Spring 1969

13. Naval Electronics Laboratory Center Report 1657, Composite OMEGA, by E. R. Swanson and E. J. Hepperley, 23 October 1969

14. Navy Electronics Laboratory Report 1305, OMEGA Lane Resolution, by E. R. Swanson, 5 August 1965

15. Navy Electronics Laboratory Technical Memorandum 1085, OMEGA Lane Resolution – Further Measurements in the Absolute Mode, by E. R. Swanson and E. J. Hepperley, 23 March 1967

16. Naval Electronics Laboratory Center TN 1472, Tests of OMEGA Receiver AN/SRN-12, by J. E. Britt, 5 March 1969

17. Naval Electronics Laboratory Center TN 1529, Accuracy Tests on TRACOR 3-599R OMEGA Receivers, by J. E. Britt, 21 August 1969

18. Naval Electronics Laboratory Center TN 1445, High Latitude D-Region, by I. J. Rothmuller, 4 November 1968

19. Naval Electronics Laboratory Center TN 1505, Auroral Zone Effects on OMEGA, by I. J. Rothmuller, 12 June 1969

20. Naval Electronics Laboratory Center TN 1429, Effect of Higher Order Modes on OMEGA Frequencies at Night, by J. A. Ferguson, 10 September 1968

21. Goddard Space Flight Center Report X521-69-346, Dual VLF Timing Capability Observed at Some Intermediate Ranges, by J. H. Roeder and M. E. Shawe, June 1969

22. Naval Electronics Laboratory Center TD 41, OMEGA Navigation System – Synchronization and Operation, 1966-1968, by E. R. Swanson and C. P. Kugel, 27 August 1968

23. Packard and Burns Electronics Publication 886B, OMEGA – A World Wide Navigational System, System Specifications and Implementation, by J. A. Pierce, 1 May 1966

24. Palmer, W., Contract NObr 92233, OMEGA Technical Note 9, Intermittent Control of Synchronization, Part I. First Order Control Systems, 24 August 1966

25. Naval Electronics Laboratory Center Report 1544, OMEGA-System Synchronization in the Absolute Mode of Operation, by C. R. Phipps, 13 March 1968

26. Navy Electronics Laboratory, OMEGA Station Standard Operating Procedure For Synchronization in the Absolute Mode, 18 February 1966

27. Naval Electronics Laboratory Center TR 1757, OMEGA Synchronization and Control, by E. R. Swanson and C. P. Kugel, 19 March 1971

28. Naval Electronics Laboratory Report, Analysis of Arctic Propagation at OMEGA Frequencies From Data Obtained at Wales, Alaska, by F. J. Rothmuller, E. R. Swanson, C. P. Kugel and J. E. Britt (In press)

29. Swanson, E. R., OMEGA, paper presented at Institute of Navigation at U.S. Coast Guard Academy, New London, Connecticut, 13-14 October 1970

TIME DISSEMINATION CAPABILITIES OF THE OMEGA SYSTEM*

Lowell Fey
Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80302

Summary

The Omega VLF navigation system affords an opportunity to disseminate time synchronization signals which could serve two classes of users: those who need precise timing and those who need time-of-day information in the form of a time code. This paper discusses precise timing use in terms of carrier pulse timing, multiple frequency techniques for carrier cycle ambiguity resolution, and Omega system capabilities in the microsecond region. For use where unattended automatic timing is needed, a time code giving second, minute, hour, and day number information could serve a variety of needs. These are described along with characteristics and constraints of such a code imposed by the existing navigation format. The paper concludes with desirable receiver characteristics and development requirements for these two timing uses.

Key Words: Multiple frequency timing; Omega time code; precise time receiver; precise timing; time code receivers; VLF timing.

It is always difficult to find out who timing customers are and what they really need. But, if they need precise timing, this implies that they are already working in a rather sophisticated technological area. They probably have some high quality equipment or money to buy it; they have made some effort to learn about possible solutions to their problems; and they tend to know when to ask help of either or both of the time dissemination agencies in this country: The U. S. Naval Observatory or the National Bureau of Standards. No present system, WWV-WWVH, WWVB included, has the worldwide, continuous availability time code potential of the Omega system. Some problems exist in adapting a time code to Omega's existing format, but these can be overcome.

An Omega Time Code

On the other hand, we know of the existence of a number of requirements for low accuracy timing which needs not only seconds but also minutes, hours, days, and perhaps years, and which is available continuously and automatically. The users for this service are harder to identify because timing, though essential, is not such a central part of their problem. This type of use generally is to time the occurrence of events which are recorded on permanent media such as strip charts or magnetic tape. Often the events occur at random times--for instance, seismic activity and other geophysical phenomena. Sometimes communication centers need to record and time all of their communication traffic. Another use is when large amounts of test data are recorded and later the results correlated in time among a number of locations. Sometimes this requires both coarse and fine timing such as for missile and satellite tracking.

For these uses it is not sufficient just to have special timing marks to identify the beginning of longer time subdivisions such as WWV's double seconds ticks which identify occurrence of the beginning of a minute. It is necessary to provide additional information to tell which minute is beginning.

Serial Time Code

A special time code which can meet this need and is applicable to Omega is the serial time code. Such a code can give the necessary timing information, repeating and updating it at a fixed rate appropriate to the timing requirements. The most well-known examples of such codes are the so-called IRIG codes. IRIG stands for Inter-Range Instrumentation Group, consisting of representatives from various missile ranges and tracking stations. This group has developed and standardized a number of codes suitable for different timing purposes.[2] All IRIG codes use pulse width encoding in binary form to convey timing information. They are transmitted either as a d-c signal level shift or as amplitude modulation of a carrier. The most well-known examples use binary coded decimal form and are not difficult to read by eye. An example, IRIG H, is shown in Fig. 1. It has a frame length of 1 minute with 1 pulse per second index markers. Wide index pulses represent a binary one and narrow pulses a binary zero. In addition, extra wide pulses are used as position identifiers to help identify the elements of the code within the code frame. Since the frame is 1 minute long the code only gives time to the nearest minute. The time to which the code refers is the beginning of the code frame.

When such a time code is recorded on a separate track along with data on magnetic tape it greatly simplifies the recovery of data. If one wishes to examine data for a specific time it is possible to locate this time with a device called a tape search unit. If a particular type of phenomenon needs to be examined, a computer unit, in conjunction with a tape search unit, can identify all events of a given class on a tape. Then these events can be reexamined in detail at low speed if desired.

* Contribution of the National Bureau of Standards, not subject to copyright.

If the time code is recorded directly on a strip chart recording along with other phenomena, the code can be decoded by eye with a little practice. An example might be a recording of phase difference between two oscillators. A time code could be added at the edge of the recording using an event marker giving time reference.

It is also possible to use time code decoders, or readers, to convert the code to an on-line digital time display when desired. Such a display will update the time reading once for each received time code frame--that is, once per minute in the examples cited of IRIG. Or, it can be provided with its own time base which is synchronized by means of the time code, but may be updated more often according to requirements.

The type of time code just discussed is the result of designing the code around a given kind of timing need. It may be generated and decoded using conventional digital logic circuit techniques. It is well suited for transmission over hard wire or high signal-to-noise radio circuits. These codes are now generally available on missile ranges and other extensive operations requiring timing. IRIG H will also become available on a 100 Hz AM subcarrier of the NBS HF stations, WWV and WWVH, starting July 1, 1971.[3] A somewhat similar though non-standard type of time code has been available from the NBS 60 kHz station, WWVB. Coverage of all these services is limited to small regions of the world and is interrupted due to changing radio propagation conditions.

We feel that a uniform world-wide 24-hour a day time code service could be provided which would benefit a much larger group of users than now can receive time codes broadcast by NBS. The Omega VLF navigation system, presently being made operational by the U. S. Navy and which will ultimately have eight transmitters located throughout the world, has the potential to meet these requirements. Therefore, we are exploring the possibilities for sharing a time code broadcast from Omega along with its navigation service broadcast. Such a code cannot be one of some six IRIG codes because of the constraints imposed by the existing Omega format, which is shown in Fig. 2. To be noted here is that the navigation frequencies, 10.2, 13.6, and 11.3 kHz are broadcast on three consecutive time segments in turn from each Omega transmitter. This is shown by dotted crosshatching in Fig. 2. (The vertical crosshatching in this figure indicates the basic unique frequency transmission slot for each station.) Each Omega time frame is 10 seconds long and contains 8 segments, thus each segment occupies 1.25 seconds on the average. In practice the segments are of slightly unequal length and are separated by 0.2 second when transmission occurs. We see that there are 5 unused segments (including unique frequency slots) which are available in turn from each transmitter. These segments have been reserved for interstation data exchange. The method used will be Frequency Shift Keying (FSK) using a different pair of unique frequencies at each station. Since there are 8 stations this makes 16 unique frequencies in all. A pair from a given station will be separated by 250 Hz. Most of these frequencies will be between 12 and 13 kHz. The required data exchange rate is sufficiently low that the Omega Planning Office and NBS have tentatively agreed that 2 minutes out of every 5 will be made available for a time code broadcast. Now we see that as a starting point in designing such a code, we have twelve 10-second Omega frames to construct a time code frame 2 minutes long. Each 10-second frame has 5 out of 8 consecutive approximately 1-second segments available. The first of these has been reserved for receiver synchronization purposes and will therefore transmit this same frequency every frame. This leaves 4 segments, appropriately the right number to transmit one decimal digit in binary form. Of the 12 time code 10-second frames in the 2 minute time code frame, the first needs to be used to identify the beginning of the time code frame and distinguish it from the 3-minute communication frame. The remaining 11 frames are available to provide a time code. A straightforward form of this code then is: minutes, 2 frames; hours, 2 frames; days, 3 frames; and years, 2 frames; leaving 2 frames unspecified. This format is shown in Fig. 3. This time code format would seem to provide the information needed by the anticipated users.

It would not be necessary for all 8 Omega transmitters to transmit a time code for worldwide reception. Eight Omega stations are planned because a navigator must be able to receive at least 3 stations simultaneously to determine his position. A time user need only receive one station to determine time. The number and location of transmitters needed to provide sufficient coverage for a time code are uncertain, but perhaps 3 would be adequate. If only 1 station is so equipped the North Dakota station would provide very good coverage of the United States.

Before discussing receiving techniques, the subject of time scales and the relation of an Omega time scale to existing time scales should be mentioned. Beginning January 1, 1972 all time and frequency broadcasts in countries supporting the International Telecommunications Union will change from the present offset frequency to exact atomic frequency and the resulting atomic time scale will be coordinated internationally by the International Bureau of Time located in Paris.[4] It will get out of step with earth time (UTC) at the rate of about 1 second per year. To prevent this, 1-second "steps" will be made in the new atomic time scale so that this scale will never depart from UTC by more than about 0.7 second. Since the steps are 1 second and the usual time signals are 1-second ticks, these steps merely amount to relabeling the ticks which are broadcast. This relabeling of seconds would not interfere with systems requiring timing which operate with a 1-second basic period. The Omega transmission format, however, has a 10 second basic period; to remain in step with the UTC system, its format would have to shift by 1/10 the basic period every time a 1-second step took place. This would contribute nothing to the navigation use; on the other hand it would most certainly jeopardize its performance. Therefore, the Omega navigation format will operate on a pure atomic scale with no steps and this will necessarily be true of any time scale it disseminates. The Omega Navigation System is already in limited operation and therefore has its own time scale now operating on the offset frequency. Changing to atomic frequency on January 1, 1972 will be disruptive enough for the system without also resetting signal phases to conform to the new international time scale. Thus, the ticks from this scale and those of the Omega atomic time scale will not coincide. This fact will not particularly affect the usefulness of an Omega time code broadcast since users will

not be concerned with differences of less than a second anyway. Precise timing users, however, will need to know the constant difference between the two scales.

Precise Timing Uses

The Omega system can be used simultaneously for precise time dissemination and time code dissemination. The multiple frequency technique for precise time recovery, essentially the same as that for navigational use, has been discussed extensively[5, 6], but will be mentioned again briefly.

For this use, it is necessary to determine the phases of two or more received frequencies. The frequencies have been arranged so that by taking differences between them, lower and lower derived frequencies can be obtained and their phases also determined. These lower frequencies then constitute timing signals with wider and wider ambiguity spacing. If time fluctuation of the phase of a given (lower) frequency is smaller than half the period of a higher frequency, a given cycle of the higher frequency can be identified with a zero crossover of the lower frequency. This method can be employed to proceed step by step from an easily identified lower frequency such as the Omega approximately 1-second segments all the way to the carrier frequency. In this way a time dissemination system with timing stability corresponding to the phase stability of the carrier may be obtained, typically a few microseconds from day to day for Omega. If propagation delay can also be determined, then the system could have an accuracy as well as a stability of a few microseconds. Portable clocks provide this possibility for such propagation delay measurement. Theoretical delay predictions may also be used, with lower timing accuracy.[7]

The number of cycle identification steps necessary to identify a carrier cycle in such a system is to first order, determined by the magnitude of the carrier phase fluctuations. As frequency differences are taken to get lower frequencies, the magnitude of the phase fluctuations, expressed in fractions of a period, are preserved.

We saw that the difference frequency fluctuations must be less than half the period of the higher frequency for cycle identification. Thus, if day-to-day averages for carrier fluctuations are 1/20 of a half period, (about 2.5 microseconds is typical of Omega) a step with a ratio of a lower to higher frequency of 20 may be used.

Since the Omega format is already fixed, only certain frequency ratios are possible. The 250 Hz separation of the unique frequencies is the lowest direct difference obtainable. With this frequency a ratio of 40.8 is required to get to the carrier frequency of 10.2 kHz in one step. This single step would probably suffice for carrier cycle identification over low fluctuation paths such as Hawaii to Boulder. Where two steps are required, the Omega format permits the ratios of 13.6 and 3, or 5.33 and 7.75, to identify carrier cycles of 10.2 kHz. The additional navigation frequencies of 13.6 kHz or 11 1/3 kHz would be used in this case.

After relating a zero crossover of the 250 Hz to a carrier cycle by the procedure just described, the next step in reducing ambiguities would be to relate the approximately 1-second Omega segment time markers to a cycle of the 250 Hz. This must be done by determining the time of arrival of these segment pulses to better than half the period of 250 Hz, that is 2 ms. Work at NBS and NELC indicates that this may be done.

Identification of the time of occurrence in seconds and minutes, etc., of Omega segment pulses may be made with the aid of some other source of coarse timing such as WWV time announcements, or it could be made from an Omega time code. In that case Omega would provide a self-contained worldwide timing capability offering a range of timing from coarse to precise.

Receivers

To make such a timing system usable, much development of a Omega time code receiver must take place. A receiver capable of giving both precise timing and of decoding the time code could consist of a number of modules. The precise timing portion must be capable of phase locking to the several carrier frequencies which are employed. Its circuits must be phase stable and a method must be provided to relate the phases of the incoming signals to the local time scale. Probably the most promising way of doing this is by a calibration technique whereby replicas of the incoming signals are generated from the local clock and then injected into the receiver front end or antenna. The receiver then becomes a comparison or null detecting device and precise knowledge of its phase shifts are unnecessary.

For determining the time of arrival of the carrier pulses with sufficient accuracy to identify a period of 250 Hz, some pulse averaging is necessary if conditions are noisy. This could be done by photographic integration or signal processing. Another way of solving this problem, described by Baltzer,[8] consists of first lowering the difference frequency from 250 Hz to 50 Hz by looking for simultaneous zero crossing of the two unique frequencies and one of the navigation frequencies. This reduction in ambiguity by a factor of five relaxes the pulse timing accuracy requirement from 2 microseconds to 10 microseconds.

For using the time code, the receiver must be capable of recognizing the presence or absence of signals during each of the 4 time code segments which occur during an Omega 10-second frame. Since signals from different Omega stations can be separated by as little as 50 Hz, the filtering problem may be severe. Once the signals are obtained from the two unique frequencies they may be recorded directly on tape or strip chart along with data from phenomena being studied. Or, if a time display is desired the logic problem of decoding them is as straightforward as with any conventional time code decoder. We believe these various developments are needed; moreover, they are well within the present state of technology and could lead to the most nearly universal timing receiver yet produced.

Conclusions

In summary, this paper proposes an opportunity for using the Omega VLF navigation system for both low and high accuracy time users. Implementation for time of day needs is suggested by means of an NBS-Omega time code (minutes, hours, days, and years) compatible with the existing Omega format. Precise timing is available through the multiple frequency technique for carrier cycle ambiguity resolution. The advantages of Omega timing are attractive, and if receiver and developmental techniques can be resolved, many timing requirements can be met throughout the world.

Refer nces

1. Richter, Henry L., Private communication, 1970.

2. IRIG, "IRIG standard time formats," Inter-Range Instrumentation Group Doc. No. 104-70, (Secretariat, Range Commanders Council, White Sands Missile Range, New Mexico 88002) August 1970.

3. Viezbicke, P. (ed.), "NBS frequency and time broadcast services--radio stations WWV, WWVH, WWVB, WWVL," Nat. Bur. Stds. (U.S.) Spec. Publ. 236, 1971 Edition (in press).

4. NBS, "UTC time scale to change in 1972," Nat. Bur. Stds. (U.S.) Tech. News Bull. CODEN:NBSTA 55(3), pp. 79, 82, March 1971.

5. Fey, L., and C. H. Looney, "A dual frequency VLF timing system," IEEE Trans. Instrumentation & Measurement, Vol. IM-15, No. 4, p. 190, December, 1966.

6. Swanson, E. R., and C. P. Kugel, "Omega VLF-timing," Naval Electronics Laboratory Center Report No. 1740, November 5, 1970.

7. Hamilton, W. and J. Jesperson, "Application of VLF theory to time dissemination," Private Communication, May 1971.

8. Baltzer, O. J., "Microsecond timekeeping by means of multiple-frequency VLF reception," Electronic Instrument Digest, Vol. 6, No. 12, p. 75, December 1970.

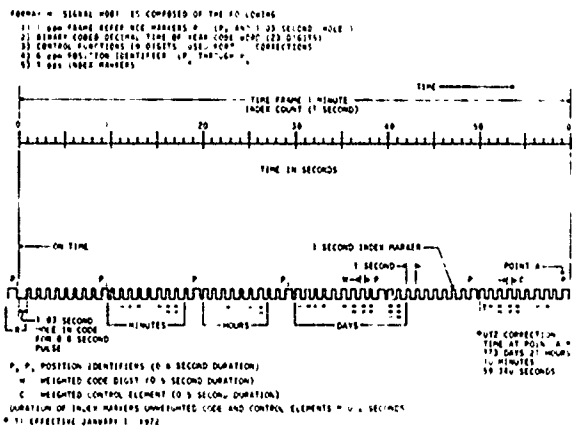


Fig. 1 - Composition of the IRIG H Format, signal H001.

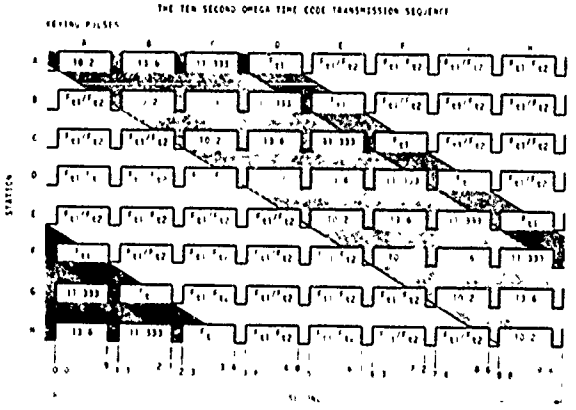


Fig. 2 - The 10 second Omega signal format.

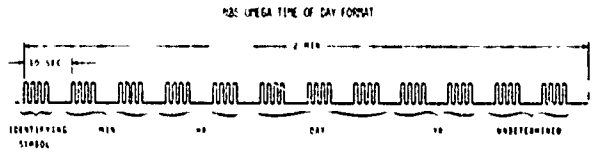


Fig. 3 - Proposed NBS-Omega time code format.

USE OF LORAN-C OVER LAND

Bernard Wieder
U. S. Department of Commerce
Office of Telecommunications
Institute for Telecommunication Sciences
Boulder, Colorado 80302

Summary

The accuracy of radio timing systems depend upon being able to predict the time delay for a signal to propagate between a transmitter (where the time is known) and a receiver (where the timing information is needed). The precision of these systems depends on the repeatibility or stability of the systems, which includes, of course, the stability of the propagation medium.

The Loran C navigation system is known for its capability to pr vide high precision for both navigation and timing. This results from the high stability of low frequency signals propagating along the ground, and because the ground wave part of the Loran C signals can be temporally separated from the more variable and less predictable sky wave part.

Correction factors are readily calculated to improve accuracy where the path is smooth and homogeneous (over sea water, for example). However, new measurements, made to assess errors in accuracy over irregular and/or inhomogeneous terrain have led to new and interesting results that must be taken into account when the Loran C system is used for timing accuracy when the propagation paths involved are over rough terrain.

1) The 100 kHz ground wave signals are slowed by the rough terrain. To take this effect into account would require an additional correction to the phase velocity equivalent in magnitude to the secondary phase correction factor calculated for smooth homogeneous land paths.

2) Measurements of the leading edge of the Loran-C pulse, which represents the pulse energy transmitted by ground wave, show a phase modulation which is highly dependent on local terrain features. Thus, phase errors of several tenths of a microsecond can occur depending on which cycle of the pulse is referenced.

3) In mountainous terrain, the amplitude of the signal can vary by as much as 15 decibels as receiver locations are changed by a few miles, while the phase discrepancy (the difference between observed and predicted values of phase) can vary by up to a microsecond from site to site.

****

Loran-C was designed as an over-the-sea radionavigation system, and the deployment of the Loran-C chains was arranged so that their principal service areas were always seaward. The Loran-C system involves pulsed 100 kHz transmissions in which a specific cycle within the pulse is identified. The bandwidth of the system (nominally 20 kHz) is great enough so that the pulse rise time is sufficiently fast to isolate the part of the signal propagating by ground wave from the parts of the signal propagated via the ionosphere. For precise position determination it is only necessary to calculate the phase delay of the very stable ground wave transmitted 100 kHz signals from the various transmitters in the chain to the receiver. If we can use a homogeneous, smooth sphere to represent the earth the delays are readily determined by calculating the phase velocity, including the diffraction corrections, for waves propagating along a geodetic of a spherical surface having a specified conductivity and dielectric constant. Since a smooth homogeneous sphere model is a good characterization of the earth for paths over sea water, Loran has worked successfully for many years and has provided very accurate positioning in the designated service areas.

In recent years, Loran-C has turned landward. In particular the installation of the station at Dana, Ind. (together with the transmitters at Cape Fear, N. C., and Jupiter Inlet, Fla.) now permits coverage for southeastern USA, and a 4 station chain provides coverage over land areas in Southeast Asia. The propagation paths are often over mountainous regions with variable conductivity. The question arises, then, what happens to Loran accuracy when the waves propagate over irregular and inhomogeneous terrain? Our laboratory, under the auspices of the U. S. A. F., is trying to answer this question.

To bring the results of our current tests into perspective it is useful to review some of the basic "smooth earth" results. The smooth earth diffraction corrections are shown in Fig. 1. These curves were calculated by Johler, Kellar and Walters[1], (also see Johler and Berry[2]) and represent the difference between the actual phase that would be measured at a given distance and the phase that would have been measured if the wave has traversed the path at the speed of light (in a medium with refractive index η = 1.000,050 to 1.000,250). The curves are parametric in earth conductivity. Generally speaking the higher the conductivity, the smaller the correction. The characteristic behavior at short distances results

from near field effects, but once beyond a couple of hundred kilometers from the transmitter the rate of change of phase is very nearly a linear function of distance and can be characterized by the slope of the curve in units of μsec/km. At low conductivities, however, there is a "turn around" in the curves. This feature is better portrayed in Fig. 2, where the slopes of the curves in Fig. 2 are shown as a function of earth conductivity. Conductivity is shown on the abscissa and increases to the left. The slope increases with decreasing conductivity, reaching a maximum for a conductivity of about 0.001 mhos/meter. For lower conductivities the slope again decreases. The physical reason for this is that at low conductivities the earth begins to react to the waves like a dielectric sphere, rather than as a conducting sphere, and the behavior of the waves changes accordingly. Note that the maximum value is about 0.0065 μs/km. For a smooth earth with .005 mhos/meter conductivity the slope is approximately 0.0042 μs/km. Thus, referenced to a .005 mhos/meter smooth homogeneous earth, the maximum difference in slope should not exceed 0.0023 μs/km, (or 2.3 μs in 1000 km).

Figure 3 shows some of the same information that appeared in Fig. 1, but it is plotted against a linear scale to better highlight the linear dependence of phase correction on distance. Also shown is the dependence of the phase correction on lapse rate, the rate of change of refractive index with height. The lapse rate can be accounted for in the calculations through the incorporation of an "effective earth's radius". The curves for $\alpha = 0.75$ correspond to the well known effective 4/3 earth radius, while the curves for $\alpha = 1$ correspond to a zero lapse rate. Thus changes in weather can change the secondary phase correction factors. The effect is small, however, (amounting to at most only a few tenths of a microsecond at reasonable distances even under extreme weather changes). For .005 mhos/meter earth, for example, the slope changes from .0042 μs/km for $\alpha = 0.75$ to .0049 μs/km for $\alpha = 1$. Thus, even though the effect is small, extreme weather variations could account for a variation of about 20 percent to 25 percent in the slope of correction factor curve. (In the data that follows the correlation of changes of slope with weather changes has not been investigated.)

When the 100 kHz ground wave signal propagate in irregular terrain, the phase and amplitude of the signals can be significantly perturbed, particularly in the vicinity of the terrain feature. Figure 4 portrays the results of some calculations by Johler and Berry[2] for the case of a 2 km high Gaussian shaped ridge. The assumed earth conductivity is .005 mhos/meter, and the results in Fig. 4 show the effect of the ridge normalized to the values that would have been calculated for a smooth earth with the same conductivity. The top of the hill is located 160 km from the transmitter. In the vicinity of the hill relatively large changes in both phase and amplitude occur. At a considerable distance from the hill, however, the fields do not quite recover and return to the values they would have had in the absence of the ridge.

Figure 5 is an artist's visualization of how these effects would impact on Loran navigation accuracy. The black lines are the lines of position that would be laid down over a smooth homogeneous earth. The white lines portray what might be expected in the presence of irregular and inhomogeneous terrain. In the mountains a considerable amount of "warp" is generated by the effect of the terrain on the signals. Far from the mountains, the black and white lines do not quite coincide because the wave fields never return to their original, unperturbed values. Also shown is a geologic fault. Calculations show that inhomogeneities in earth conductivity cause perturbations quite similar to those created by terrain irregularities. Since terrain irregularities and conductivity inhomogeneities usually occur together it is hard to experimentally separate out the effects of one from the other.

Under the auspices of the USAF the Institute for Telecommunication Sciences of the Office of Telecommunications, undertook experiments to test these effects. They were undertaken in two stages. The first stage was designed to evaluate the long range effects, and the second stage, still in progress, was to examine the local effects. For the first stage a C-131 aircraft was equipped with a specially modified ARN-85 navigation set which permitted the comparison of the phase of the 100 kHz signal received from a given transmitter with the phase of a 100 kHz signal aboard the aircraft that was derived from a rubidium standard. In addition, the aircraft was equipped with a mapping quality aerial camera so that the actual position of the aircraft could be determined accurately. Figure 6 shows the three flight profiles flown during the tests. All profiles were repeated several times, usually at a height of 7000 ft above sea level. The E-W path over N. Carolina-Tennessee and the N-S path over Pennsylvania were along geodetic lines intersecting the transmitting station at Cape Fear, N. C. The E-W path over W. Virginia-Maryland was along

a geodetic intersecting the station at Dana, Ind.

Figures 7, 8, and 9 show typical results from the flights. The abscissa in each case shows the distance from the referenced transmitter (Cape Fear for the N. Carolina-Tennessee and the Pennsylvania flight profiles, Dana for the Maryland-W. Virginia flight profile). The uppermost curve shows the terrain altitude along the geodetic under the flight profile. The straight line in the middle curve corresponds to the specified geodetic and the line that wanders about it shows the actual flight plan. The plans for the experiment included a requirement that the aircraft stay within 1/4 mile of the specified flight profile, but it was only rarely that we succeeded in achieving this goal. In most flights, the airplane was sometimes more than a mile from the designated path so that it was not possible to correlate small scale variations on the data with identifiable land features. The lowermost curve in each figure shows the phase change, normalized to the smooth homogeneous earth model values for an earth with 0.005 mhos/meter conductivity. The slope for the smooth earth case is 4.18 μs/Megameter (a Megameter equals 1000 km). The slopes observed for the N. Carolina-Tennessee, Pennsylvania, and the Maryland-W. Virginia flights, obtained by least squares fit, are in excess of this by 4.61 μs/Mm, 4.21 μs/Mm, and 4.53 μs/Mm, respectively. These values far exceed the values that could be predicted for any value of earth conductivity using smooth earth theory. Thus, it appears that the rough terrain does indeed slow the waves. Of course, these data were taken at an altitude of 7000 ft above sea level. There is some indication in the data that there is an altitude effect, but nevertheless, the effect is real and undoubtedly accounts for much of the time difference discrepancy observed in Loran lines of position in inland service areas.

To examine local terrain effects we chose an extremely rugged area of the Appalachian range. The area is near Asheville, N. C. and includes Mt. Mitchell, the highest point in eastern U. S. A. The instrumentation originally on the C-131 is now aboard an Air Force helicopter and we will intensively cover a relatively small area in order to map the scale and magnitude of the local anomolies In support of this effort, we made ground-based measurements using Loran timing receivers with a cesium standard for a reference.

Figure 10 shows the area in which the ground-based data were taken. The black dots identify the 15 locations used. These are numbered in the order that we first went to each site. We shall use the numbers to identify the sites in the figures that follow. Note that the altitude in the area varies from about 2200 ft to about 6800 ft above sea level and that the maximum horizontal distance involved is only about 25 km.

Figure 11 shows the tracing of the front end of the Loran pulse produced by the timing receiver. The short horizontal section indicates the position of the "sampling point", i.e., the zero crossing in the pulse locked onto by the receiver to provide a phase reference. The receiver also has an amplitude strobe which occurs 1/4 cycle prior to the sampling point. The chart on which the pulse was drawn is driven by a inverter but one second time ticks derived from the receiver 100 kHz were recorded on the chart and used to insure that the chart speed was constant.

When the zero crossings (both positive going and negative going) of these tracings were scaled and plotted on a chart with a common scale, we were surprised to find that the length of the first 3 cycles of the pulse varied from site to site. Figures 12, 13, and 14 show the results for the three signals from Cape Fear, Dana, and Jupiter. The vertical bars indicate the position of the zero crossings. The scales were normalized to be equivalent to 30 μs for the longest trace. The results were sorted longest on top, shortest on bottom, so that the trend appears clearly. Note that the order of the site numbers differ between figures. Thus, the effect is different at a given site for each of the three signals and cannot be simply related to characteristics of the sites alone. This phase modulation on the pulses is undoubtedly due to radio energy that is scattered by the rough terrain which interferes with the signals that come directly from the transmitter.

This interpretation becomes more credible if we look at the amplitude and phase characteristics of the signals. Figures 15 and 16 show the amplitude and phase of the Dana signals as a function of site position. Here the sites are ordering in sequence according to their distance from the Dana transmitter, but normalized so that the closest site is set to zero. In Fig. 15 the open circles connected by the light lines show the variation in amplitude. There is an excursion of almost 30 dB in signal strength at different sites within a distance interval over which smooth earth theory would predict less than 1 dB change. The squares connected by heavy lines indicate the elevation of the various sites. There appears to be a correlation between amplitude and elevation, and statistical evaluation indicates that for these 15 data points the

correlation is significant.

Figure 16 shows the variation in phase difference from site to site. The quantity (Δ phase) plotted is related to the difference between observed phase and the phase predicted on the basis of smooth earth theory with 0.005 mhos/meter conductivity. It differs from observed less predicted phase by an arbitrary constant so that the variation in Δ phase from site to site is of significance here. Note that Δ phase varies by as much as 3 μs over relatively small incremental distance. Contrary to the amplitude results, this variation is not significantly correlated with elevation.

These data indicate that without proper calibration the phase of the Loran signals propagating over land can differ from predicted values by a few microseconds and that the discrepancy can be attributed to rough and/or inhomogeneous terrain. The discrepancies can be separated into those attributible to long range propagation effects, and those ascribable to local effects. The former can readily be corrected through simple but systematic area calibration. The latter is more difficult. But the phase and amplitude variations shown in the last two figures can be viewed from the point of view of holography, so that with adequate data the wave-fronts can be reconstructed and the source of scattering located. With this information we should be able to predict propagation effects much more accurately than we can now, and hopefully provide time via Loran with microsecond, or submicrosecond, accuracy.

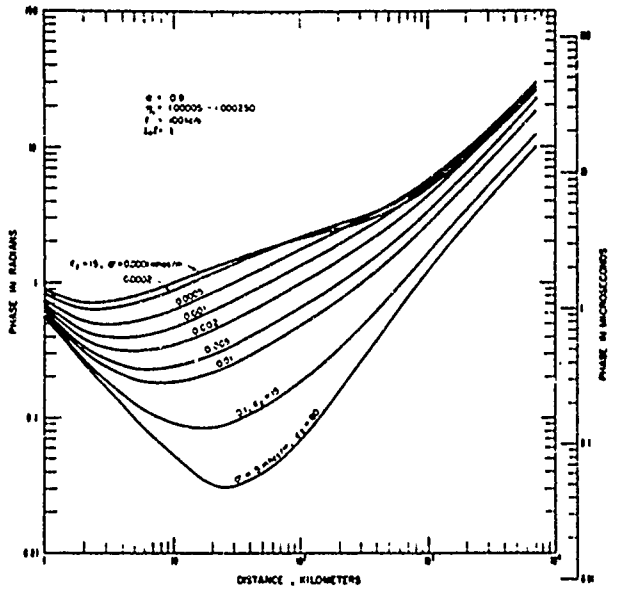
Acknowledgements

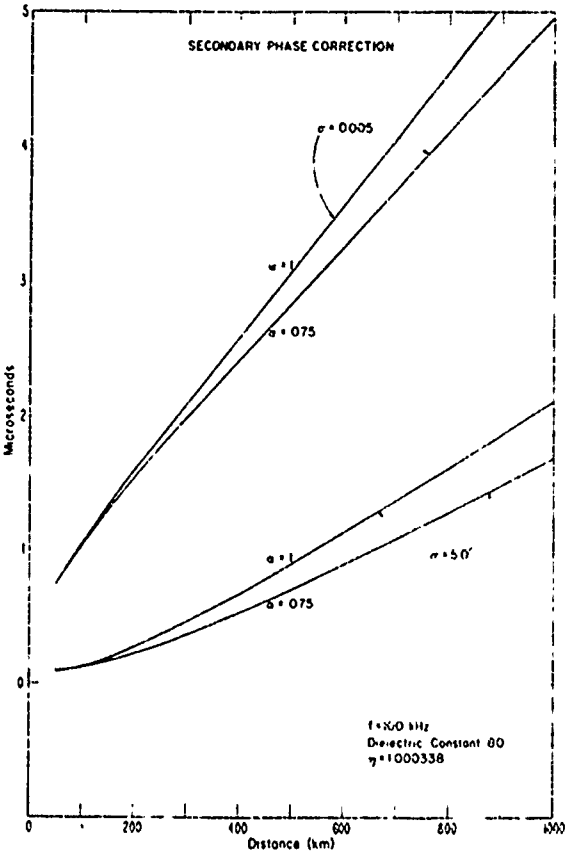
The results given in this paper stem from work that was done under the auspices of the USAF through the Loran Engineering Office, Aeronautical Systems Division, Wright Patterson Air Force Base. I have drawn heavily on the work of J. S. Washburn, R. H. Espeland, and J. R. Johler, all of whom have participated in the project. The airborne data was taken by the Air Force under the program direction of W. Augustine and A. Mewborn.

References

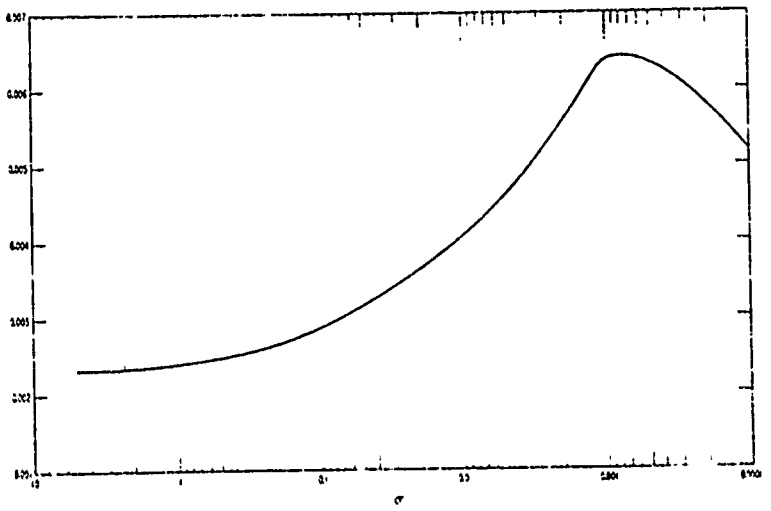
1. Johler, J. R., W. J. Kellar, and L. C. Walters, (1956), Phase of the low radio frequency ground wave, NBS Circular 573 (U. S. Government Printing Office, Washington, D.C. 20402).

2. Johler, J. R., and L. A. Berry (1967), Loran-D phase corrections over inhomogeneous, irregular terrain, ESSA Tech. Rept. IER 59-ITSA 56 (U. S. Government Printing Office, Washington, D.C. 20402).

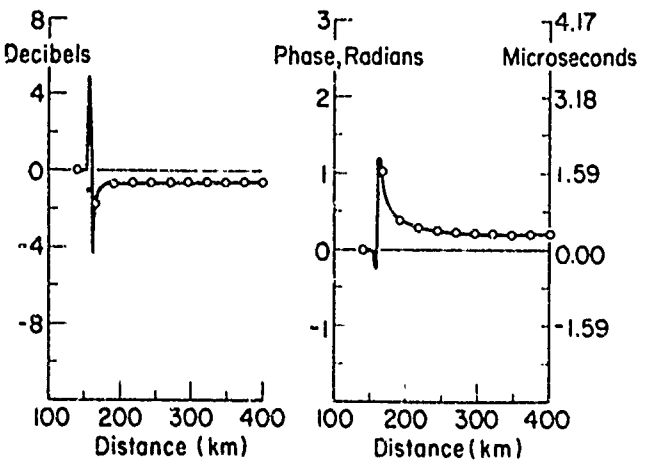


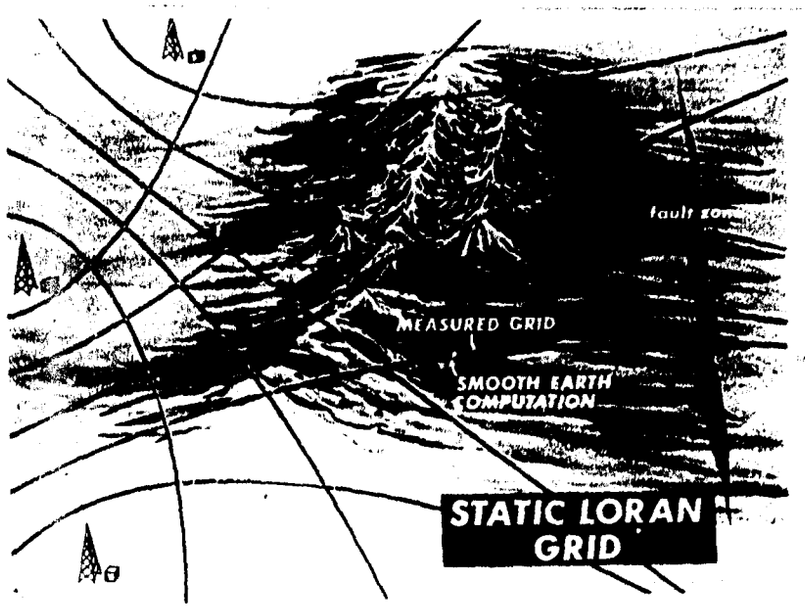


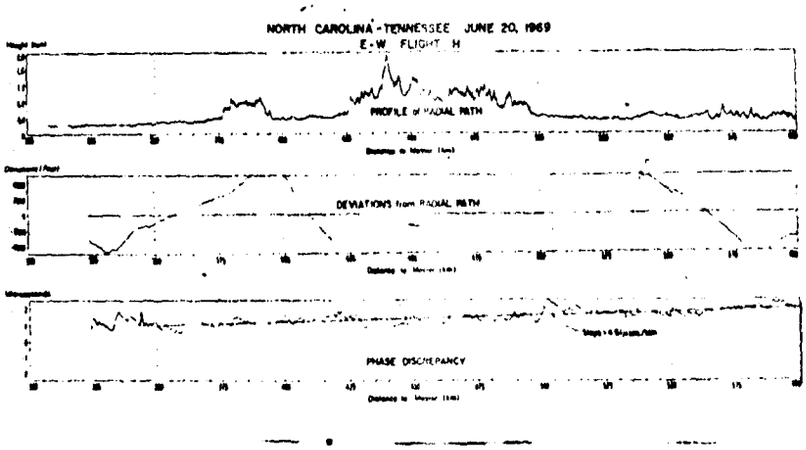
AVERAGE μsec/km PHASE CORRECTION, 1000 2000km INTERVAL

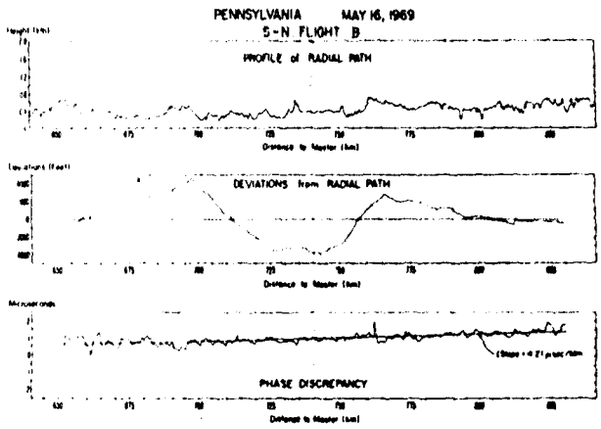


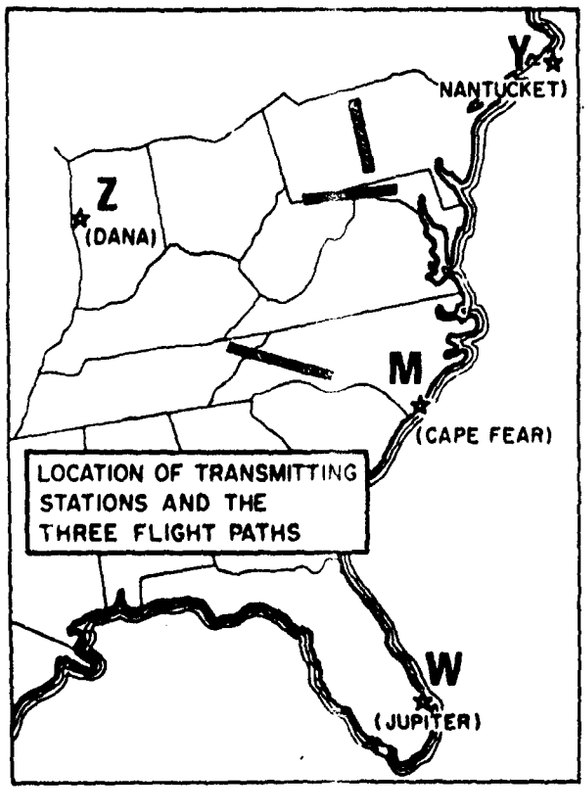
AMPLITUDE AND PHASE PERTURBATIONS IN PROPAGATION OVER A 2 KM HIGH GAUSSIAN SHAPED RIDGE

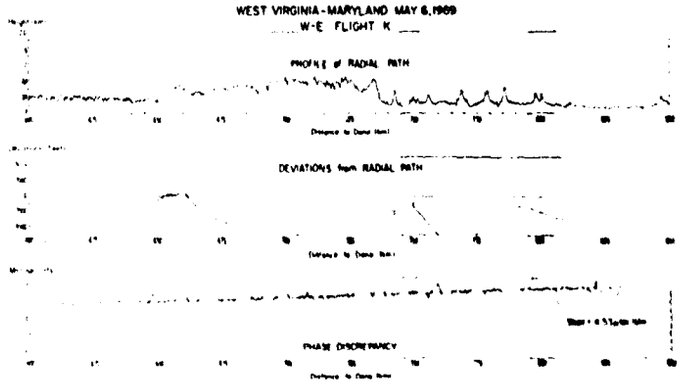



fault zone
MEASURED GRID
SMOOTH EARTH COMPUTATION
STATIC LORAN GRID

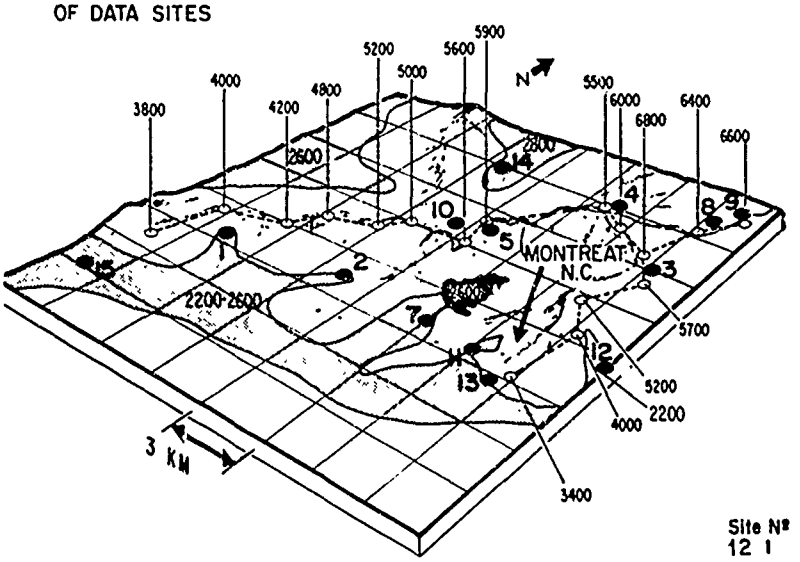

NORTH CAROLINA - TENNESSEE JUNE 20, 1969
E-W FLIGHT H
PROFILE of RADIAL PATH
DEVIATIONS from RADIAL PATH
PHASE DISCREPANCY

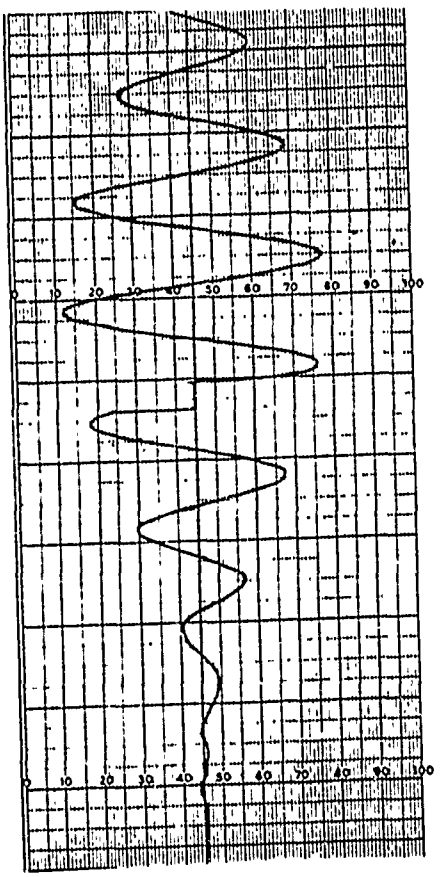

PENNSYLVANIA MAY 16, 1969
S-N FLIGHT B
PROFILE of RADIAL PATH
DEVIATIONS from RADIAL PATH
PHASE DISCREPANCY

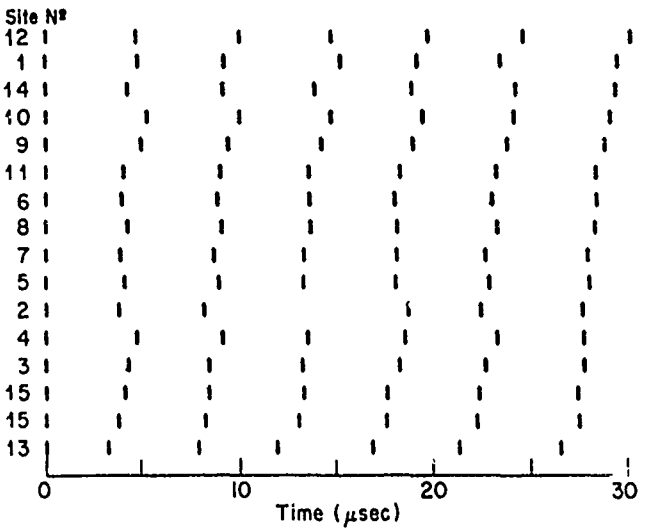

Y
NANTUCKET)
Z
(DANA)
M
(CAPE FEAR)
W
(JUPITER)
LOCATION OF TRANSMITTING STATIONS AND THE THREE FLIGHT PATHS


WEST VIRGINIA - MARYLAND MAY 6, 1969
W-E FLIGHT K
PROFILE of RADIAL PATH
DEVIATIONS from RADIAL PATH
PHASE DISCREPANCY

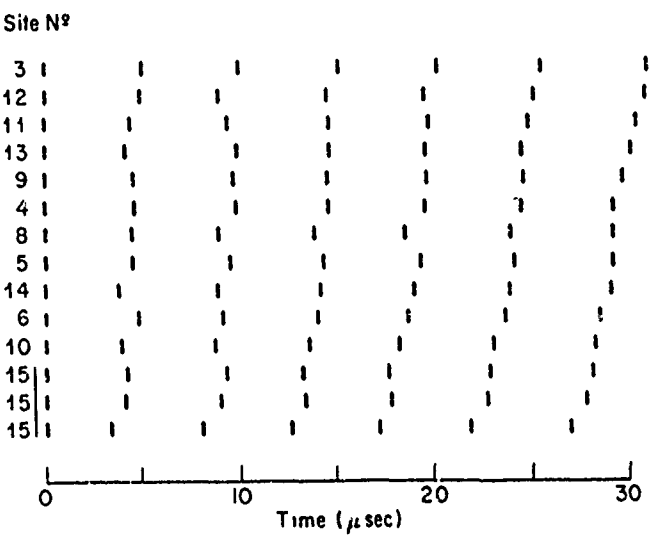
DETAILED MEASUREMENTS AREA SHOWING PROMINENT GEOPHYSICAL FEATURES AND LOCATIONS OF DATA SITES



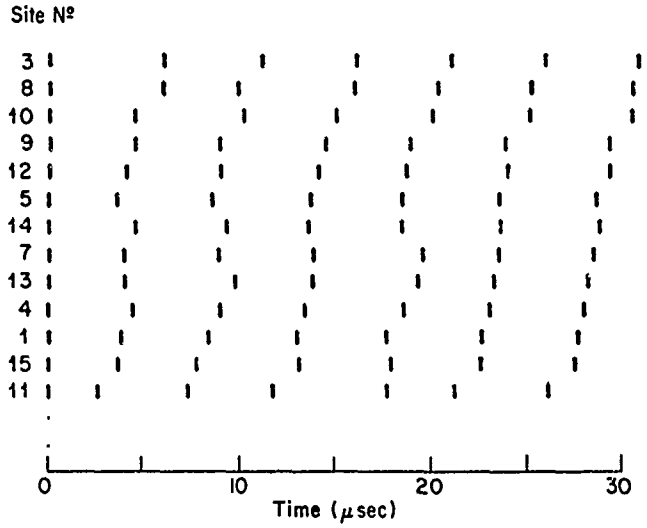




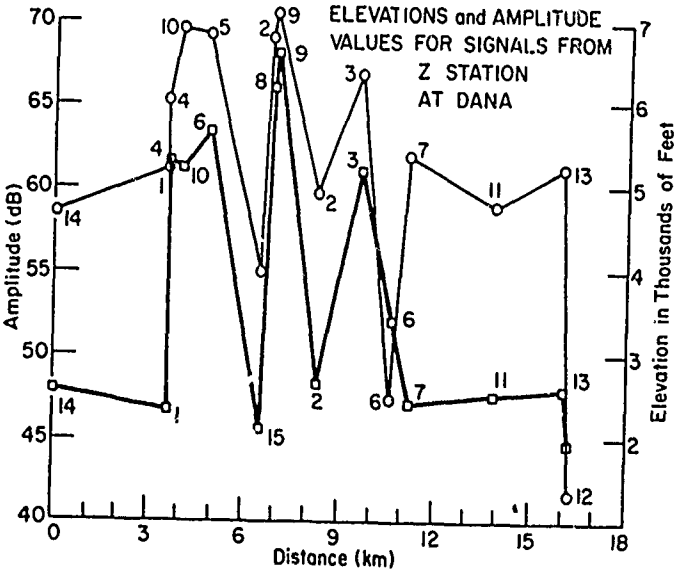
ZERO-CROSSINGS FOR SIGNALS MEASURED NEAR ASHVILLE, N.C. FROM THE M STATION (CAPE FEAR)

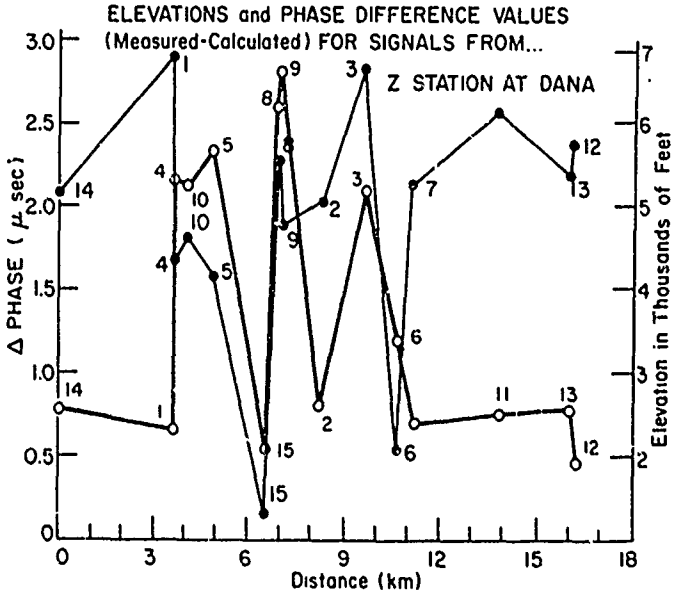


ZERO-CROSSINGS FOR SIGNALS MEASURED NEAR ASHVILLE, N.C FROM THE Z STATION (DANA)



ZERO-CROSSINGS FOR SIGNALS MEASURED NEAR ASHVILLE, N.C.
FROM THE W STATION (JUPITER)





ONE-WAY TIME DISSEMINATION FROM LOW ALTITUDE SATELLITES

L. J. Rueger
The Johns Hopkins University
Applied Physics Laboratory
Silver Spring, Maryland

Summary

Geodetic and navigation satellites operate in coordinate and time systems that permit accurate dissemination of time to any number of ground based sites. Time Recovery from these satellites requires the calibration of the satellite time scale, the calibration of the passive ground based receiver systems and the calculation of the propagation time for the signals to travel from the satellite to the receivers. This paper will discuss technology and capability for time synchronization to simultaneously observed events, and for precision time dissemination from an orbiting clock to any passive receiving site on the globe.

Introduction

The use of radio signals to disseminate time dates back to 1905 when the Navy initiated time broadcasts to calibrate chronometers in ships. Since that time, various radio links have been provided from terrestrial stations for both military and civilian services. As the accuracy of these time signals available at the transmitting stations improved, the errors associated with the propagation medium became the dominate error for paths in excess of several hundred miles.

The advent of artificial satellites with active transmitters opened new technology to reduce propagation errors, by permitting line-of-sight propagation on very high frequency links.[1] The properties of the propagation medium using the satellite-to-ground radio link contribute less than 1 microsecond to the measured error.[2] Other error sources, such as uncertainties in the signal path lengths, station locations, and bandwidth/power relations become the dominant effects.

In this paper, I will discuss how these effects are handled and how stations beyond simultaneous line-of-sight to a single satellite can be synchronized in time - up to any global separation.[3]

System Configuration

The systems to be considered are a family of low altitude satellites which provide one-way time dissemination to passive receiving stations by carrying precision clocks. Several satellite programs have provided this type of time signals: ANNA, GEOS, DODGE, TIMATION, AND TRANSIT. Of these, all are in the category of low altitude (<1000 n. mi. mean altitude) except DODGE which is slightly below synchronous altitudes. Each system provides time signals from the precision oscillator that are used to modulate the RF carriers.

The geometry of this system is shown in Figure 1. In order to synchronize times t_1, t_2, and frequencies f_1, f_2 at two ground sites via signals received from the spacecraft's time system t_3, f_3, it is essential to know the positions of Site 1, Site 2 and the spacecraft. The distances

Site #1-to-Satellite = $\sqrt{(x_1-x_3)^2+(y_1-y_3)^2+(z_1-z_3)^2}$

Site #2-to-Satellite = $\sqrt{(x_2-x_3)^2+(y_2-y_3)^2+(z_2-z_3)^2}$

must be known quantities to account for the propagation time from the spacecraft to each site. In the case of fixed sites and geodetic or navigational satellites, these distances are a part of the information available from either the survey or navigation computations performed in the normal use of the satellite.

Obviously, a specific event in the satellite that can be observed at both Site 1 and Site 2 is the most direct process for synchronizing sites 1 and 2. For satellites at 600 n. mi. altitudes above the earth, this mode of operation applies for sites separated by less than 3500 n. mi.

At greater separations of ground sites, the satellite is not simultaneously visible and therefore events observable at each site must be related to the time system carried in the satellite as observed by a fixed station on the ground. Relativistic shifts of the time base in the spacecraft are very small if the orbits are circular.

The computation of the radio path length is complicated because the spacecraft position is best described as simple Kepler motion in inertial space while the positions of the receiving sites are best described in earth-fixed coordinates rotating in inertial space. Fortunately, these distances are determined in the navigation and geodetic survey calculations and are readily available for time delay corrections.

Epoch Markers

Each satellite time system provides an unambiguous event every few minutes to resolve timing ambiguities associated with the high frequency modulations needed for precise time resolution. For example, the TRANSIT program generates a coarse time signal every 2 minutes, and a high resolution time signal every 10 milliseconds. The high resolution signal is a phase modulation with a 90% rise time of 1 microsecond. All time signals including the RF carriers are derived from the same stable reference oscillator.

The Transit system requires a self-consistent time and frequency but there are operational advantages to relating this time system to other global timing systems, namely, UTC. For the functional self-consistency needs of the present doppler navigation program, the time must be maintained to better than 2 milliseconds throughout the system. The geodetic work, which supports and establishes the geodetic reference structure to the navigation system, can justify time being maintained to better than 200 microseconds.

Now that you understand these basic scientific and operational program requirements, I want to show you the timing provided in the NAVSAT Program.

The time base or clock rate is generated in the satellite by an integer division of the primary oscillator frequency. Normalization to UTC time is accomplished by deletion of 10 microsecond steps in the clock rate to compensate for the predicted crystal oscillator aging determined by prior measurements. The deletion of 10 microsecond steps is controlled from the ground by observing the accumulation of clock error and programming a pattern of 10 microsecond adjustments to occur throughout the day to steer the clock error to zero This pattern is determined and reset once per day in each satellite.

In Figure 2 is shown the epoch error relative to UTC as measured for the constellation of 5 satellites for January, February 1971. The error is rarely greater than 50 μsec. The satellites of the present design could be held to about 10 μsec. error by changes in the prediction and control procedures used to operate the system.

Time Transfer

The data shown in Figure 3 indicates the error for time transfers for two widely separated passive stations using non-directional antennas. The error is made up of ground equipment error, satellite clock error in 107 minutes and computed propagation delay error. The calibration of the satellite time scale to UTC requires (a) the measurement of signals received at a ground station whose clocks are in step with UTC, and (b) computing or measuring the propagation time required for signals to travel from the satellite to the point in the instrumentation where the measurement is made. The timing signals are transmitted at the satellite in synchronism with UTC. Every user must determine the propagation time from the satellite through his receiving equipment. The calibration ground station is put in step with UTC by Naval Observatory time transfers for epoch setting and is maintained in step by monitoring of a Naval Observatory controlled reference radio signal.

In order to time synchronize globally-remote stations, each station must have a clock or time system capable of holding below an allowable error over the interval between opportunities to get calibrated satellite signals. For low-altitude, polar satellites, global coverage is possible every 12 hours from a single satellite. However, with 5 satellites the typical interval is two hours. The remote site can reduce the probable time synchronization errors by using multiple observations. The multiple observations can be made in several ways, for example; several times during a single pass of the satellite, or from successive passes from various satellites or from successive passes from the same satellite but using a long span of time. The systematic errors and biases vary with each of these observational methods.

Table I is a list of factors that must be taken into account in effecting a time transfer from one terrestrial station to another via low altitude satellites. The most significant factor in the satellite instrumentation is the drift in the reference oscillator that drives the satellite clocking system. The most effort required for each time transfer computation is the determination of the propagation path delay that goes with each epoch measurement. In geodetic and navigation satellite programs, these computations can be made with great precision. The most significant factor in the receiver performance is calibration of the receiver delays and minimization of signal jitter due to noise.

In geodetic and navigation satellite programs, the transmitter power has been so low that narrow band phase-locked receivers have been required to obtain reliable signal reception. This factor has dictated relatively low modulation frequencies and consequently limited the epoch resolution in the receiving instrumentation.

Time Disseminating Satellites

Design choices that were made in some of the low altitude satellite programs shall now be identified.

The ANNA satellite was part of a program to evaluate simultaneously several navigation techniques: range rate doppler, microwave transponder ranging, and optical angle tracking on a star background. The ANNA satellite (1962-65) marker epochs were coherent with the RF carriers and no attempt was made to synchronize to UTC.

Epoch was identified by phase reversal of the square wave modulating function near the end of the timing burst. Figure 4 shows time difference data taken from this satellite using high gain antennas. The data is shown without and with correction for slant range. The corrected data shows a measurement resolution of about 100 microsec.

Table I
Factors to be Included in Satellite Time Transfer Analysis

1. Satellite Subsystem
   a) Epoch Marker Generator
      *Reference Oscillator Drift
      Frequency and Time Translation Circuits
      Modulator Time Jitter
   b) Transmitter Time Delay Stability.
2. Radio Path and Satellite-to-User Geometry
   *a) Computation of Satellite Position vs. Time in Earth Fixed Coordinates.
   b) Determination of Receiver Position vs. Time in Earth Fixed Coordinates.
   c) Corr ction for Ionospheric Refraction of the Radio Wave Propagation Path.
   d) Correction for Tropospheric Refraction of the Radio Wave Propagation Path.
3. Terrestrial Receiving Subsystems
   *a) Calibration of Receiver Delays.
   b) Jitter in Signal from Noise.
   c) Resolution Uncertainty in Epoch Decision Logic.
   d) Dynamic Signal Tracking Bias.
   e) Clock Comparison Circuit Error.
   f) Station Clock Drift.

* Most significant.

The GEOS satellite (Figure 5) was part of a NASA-sponsored geodetic program and still provides timing signals for synchronization of remote tracking stations. You can see the same types of systems as carried on the ANNA; doppler, microwave transponder, flashing lights and laser reflectors. Timing was provided by phase modulation of the carriers, as shown in Figure 6. The epoch is recovered by detection of the sudden phase reversal of the modulation.

The GEOS time system has a time normalization with a capability to hold 1 minute epoch markers to within ± 10 microseconds of a prescribed system reference. Timing of remote stations to better than 50 microseconds of UTC has been achieved with this system.

The TIMATION satellite program is a Navy-sponsored low altitude system using ranging for navigation. The timing performance of this system is not available for dissemination to this conference.

The NAVSAT satellite program is a Navy-sponsored operational navigation system that is used by submarines and ships in providing all-weather global high accuracy navigation fixes. As a secondary function each satellite transmits timing epochs which are normalized in scale and are calibrated to UTC by a ground support network. This is an operational system readily available on a global basis now and will be available for some considerable time into the future, so I shall describe in more detail the system characteristics and the instrumentation involved in time transfer.

Orbit

The satellites are in circular polar orbits 600 n. miles above the earth. Nominal spacing of the orbit planes of 45° at the equator requires 4 satellites in service. Precession of the orbital planes with time has changed this nominal spacing, and 5 satellites have been put into service to increase the availability of global access for navigation.

Configuration

Figure 7 shows one of the present operational satellites. It is much smaller than the others described, and can be launched on one of the smallest launch vehicles. It weighs about 112 pounds deployed in orbit. Power is provided by solar cells and rechargable storage batteries. Gravity stabilization permits the use of a directional circularly polarized antenna.

Block Diagram

Figure 8 is a block diagram of the satellite. An oven controlled, 5th overtone crystal oscillator drives a multiplier chain to provide coherently related RF carriers at 150 and 400 MHz. The same source is divided down to provide timing signals at even two minute marks.

Orbital Data

The RF carriers are modulated with digitally encoded data describing the orbital path of the satellite as a function of time. This data is provided from ground support facilities and transmitted to the satellite memory storage for use as it proceeds in its orbit. The same ground facilities recover the satellite transmitted signals and measure the doppler shift, the oscillator frequency and satellite epoch. This data is used to predict the future orbital path of the satellite and to program the satellite clock epoch.

The predicted orbital path is strongly dependent on the earth's gravitational field. Over the last 10 years, extensive efforts have been made to measure this function by observing satellites from fixed stations well spaced around the globe.

One of the important tasks related to this set of fixed tracking stations is station time synchronization to a common epoch.

Tracking Station Timing

Satellite time signals have been used to measure the epoch variations in the network station clocks. The RMS timing residuals for all the stations in the network for April 1970 are shown in Figure 9. Only two of these 330 and 111 have atomic quality

reference standards, and the superb performance of these stations is clearly seen. This data differs from the early histograms shown because the station clock error is the dominant effect seen here and all satellite clock drift has been removed.

Timing Receiver

The operators in one of these stations built a special purpose receiver primarily for recovering time signals from the NAVSAT satellites and ignoring the usual doppler data needed for navigation. This instrument can provide the raw data for epoch setting a remote station clock, but must have propagation path lengths provided from an outside source. Figure 10 identifies the engineering design parameters. These units are commercially available from a small private company.

Orbiting Oscillator Stability

The stability of the oscillator in one of the orbiting NAVSAT satellites is shown in Figure 11. It changed frequency by about 1×10^{-9} in 100 days service. The long term drift rate is about 1 part in 10^{11} per day. The short term (several seconds) stability is also about 1 part in 10^{11}.

Incremental Phase Shifter

A program has been authorized to develop another series of satellites to follow the present NAVSAT. One of the items planned for this series is an improved way of normalizing time to the UTC standard. A device called an incremental phase shifter has been developed to take the output of a frequency drifting oscillator and normalize the frequency to a prescribed value. The frequency transfer function is exactly predictable by digital logic circuitry whose function can be programmed by ground station radio commands. The range of adjustment will cover the 5 year aging range of a crystal oscillator and progresses smoothly over the entire range in 8×10^{-13} steps.

The logical block diagram of the device is shown in Figure 12. The basic principle employed is to introduce the 5 MHz signal into a delay line system of 200 - 1 nanosecond steps, and take the output of a digitally driven commutator switch that cyclically scans the delay line. The number of parts required are reduced by building the 200 steps from two units--one with 8 steps of 25 nanoseconds each, and the other with 25 steps of 1 nanosecond each. In this way 31 switches (25+8) does the job of 200. The switch arm is advanced to each position at a rate determined by ground command. The ground command sets the number of 5 MHz cycles of the satellite oscillator that must be generated between each step advance of the switch.

Based on the measured satellite oscillator aging rate, a predicted rate can be programmed for correction including 2nd and 3rd order rates of change.

Future

This technology opens the way for providing nearly atomic reference oscillator stability to orbiting crystal oscillators and retaining the long service characteristic of crystal oscillators

When the day comes to orbit atomic standards, this device provides the vernier frequency adjustment to set the time to any convenient system rate and to normalize out the differences between standards of the same generic design.

For the immediate future, this design makes possible time transmissions with better than 1 nanosecond epoch settability.

References

1. TG-459, "Precision Time and Frequency Applications in the TRANSIT and ANNA Satellite Programs", L.J. Rueger, Dec. 1962, APL/JHU Technical Document.
2. TG-1086, "Global Timing Systems of Nanosecond Accuracy Using Satellite References" E. F. Osborne, Oct. 1969, APL/JHU Technical Document.
3. "Proceedings from Precise Time and Time Interval Strategic Planning Meeting", Naval Observatory Dec 1970 Vol. 1.

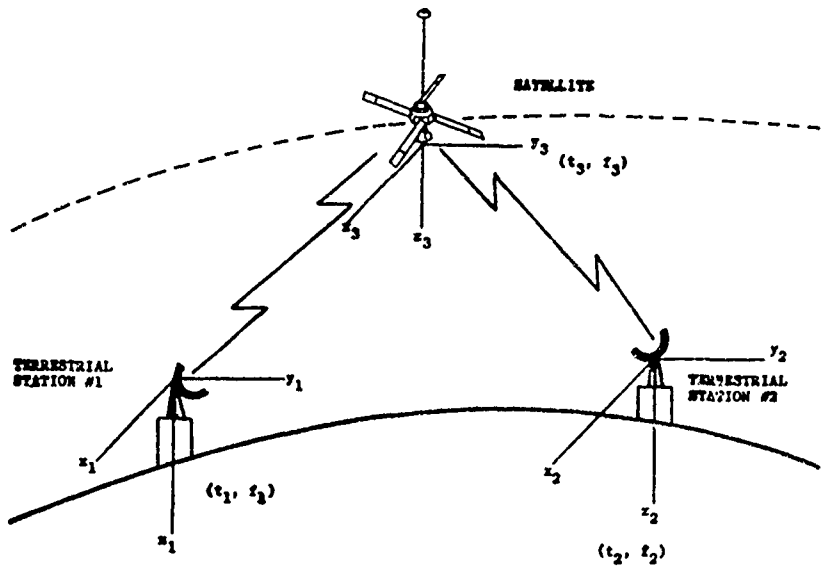


FIGURE 1 - SATELLITE TIME TRANSFER GEOMETRY

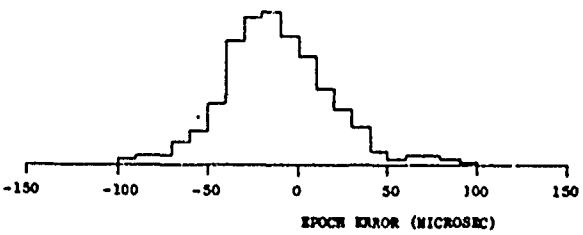


FIGURE 2 - HISTOGRAM OF NAVSAT SATELLITE EPOCH ERROR
FOR JANUARY-FEBRUARY 1971
(538 INJECTION STATION MEASUREMENTS)

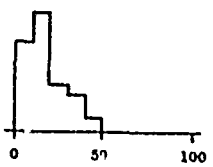


FIGURE 3 - HISTOGRAM OF 65 SATELLITE TIME TRANSFERS
BY STATION 110

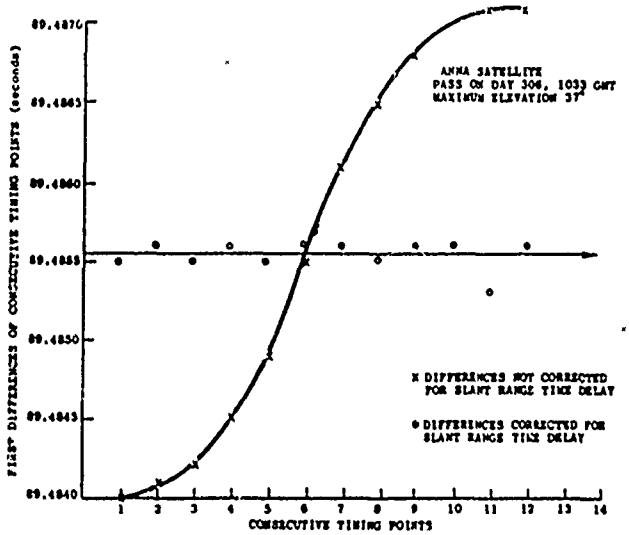
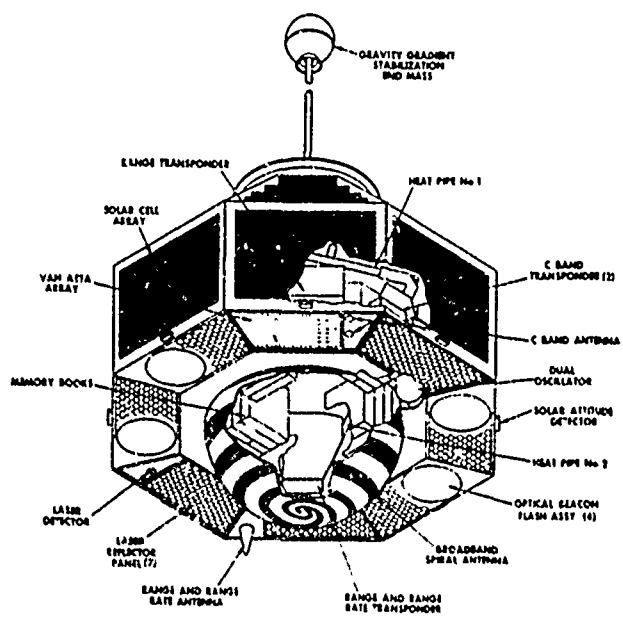


Fig. 4 ANNA TIMING POINTS TAKEN AT STATION 091

GEOS B SPACECRAFT

FIGURE 5



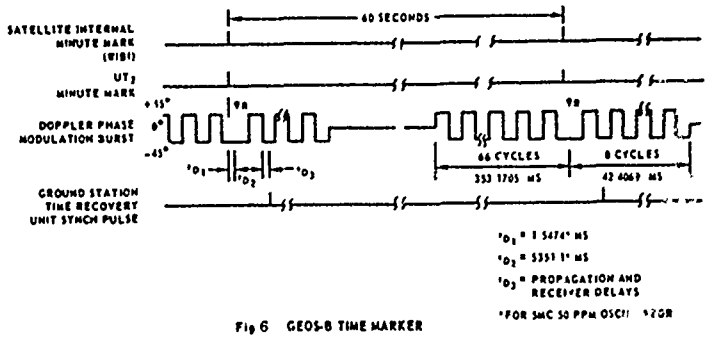
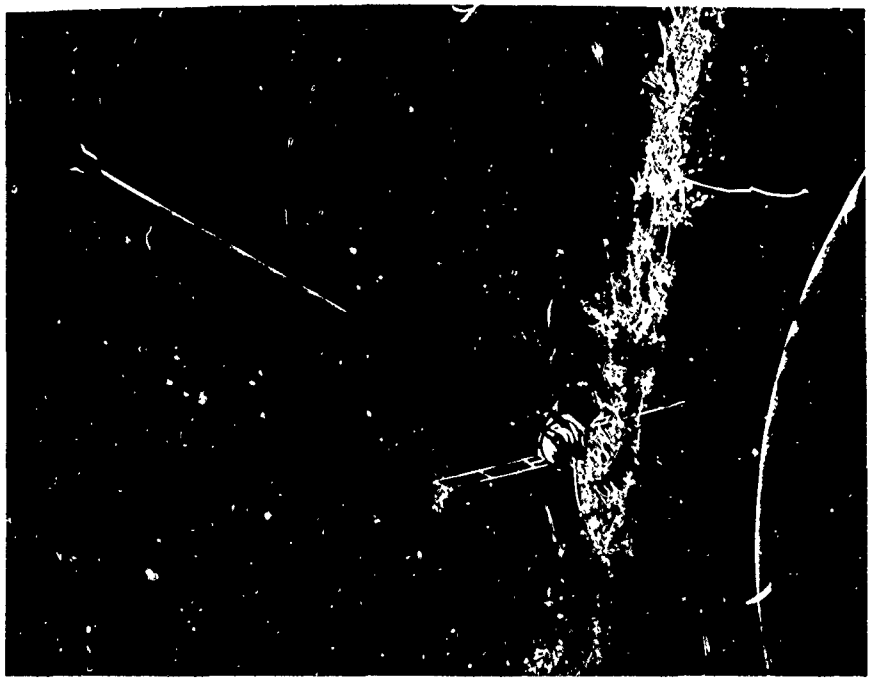


Fig 6 GEOS-B TIME MARKER



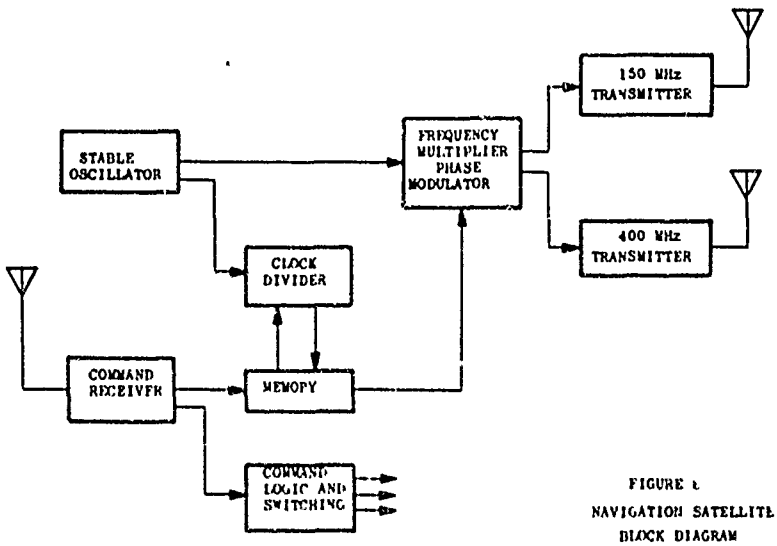


FIGURE 6
NAVIGATION SATELLITE
BLOCK DIAGRAM

FIGURE 9

DISTRIBUTION OF APL TIMING RESIDUAL RMS FOR
SATELLITE 1967-92A, 1-30 APRIL 1970

Station	RMS (Milliseconds) .01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.11	.12	.13	.14	.15*	Total
008	2	1		3	3											8
013	1	2	2	4	1											10
014	10	7	1	1												19
018	5	11	14	17	8	5	2	1		2					1	66
019	5	20	21	11	4	1	1									63
020	4	4	5		1											14
103	7	6	5		1											19
106	5	1	7	4	4	2										23
111	20	3														23
112	1	4	12	5		1										23
115	1	6	7		2											16
117	2	6	5	2	1			1								17
121	3	5	3	5	2											18
330	16															16
700			1													1
733	12	1														13
765	1	2	5	4	2	3		1								18
766	1	1	4		2											8
893			2	1		1	3	2								9
896	12	19	8	6												45
898		4	3	6	6	4	1	1		1						26
911	5	2														7

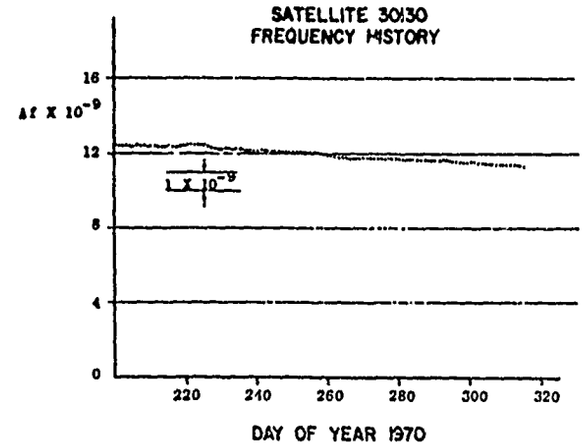
*0.15 or over.

FIGURE 10

NAVIGATION SATELLITE TIME RECOVERY RECEIVER

1. FIRST MODEL: 8 μ second rms @ -115 dbm SIGNAL/NOISE = 6 db
2. SECOND MODEL: 5 μ second rms @ -115 dbm SIGNAL/NOISE = 3 db
3. OPERATING FREQUENCY. 400 mHz - 80 PPM (±2 X 10^{-5} DOPPLER)
4. INFORMATION BANDWIDTH OF PHASE LOCKED LOOP: 15 Hz
5. OUTPUT: PULSE - 10 VOLTS WITH 0.2 μ second RISE TIME EVERY 2 MINUTES
6. SIZE: 3½" PANEL SPACE; 19" RELAY RACK WIDTH; 15 POUNDS
7. POWER: 10 WATTS @ 60 ~
8. COST: ABOUT $ 2500 IN QUANTITIES OF TEN

FIGURE 11



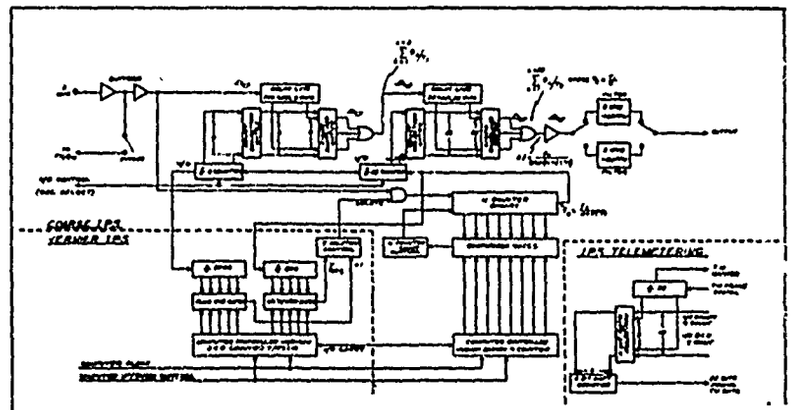


Figure 12 - Incremental Phase Shifter System Block Diagram

TIME TRANSFER BY DEFENSE COMMUNICATIONS SATELLITE

J. A. Murray, D. L. Pritt, L. W. Blocker, W. E. Leavitt and P. M. Hooten
Naval Research Laboratory
Washington, D. C.

and

W. D. Goring
Naval Electronic Systems Command
Washington, D. C.

Summary

Time pulses are effectively passed from one Defense Communications satellite terminal to the other as specified points on the high-speed pseudo-random codes used by the communications modems at the stations. If the specified points of the codes used in both directions of transmission occur at approximately the same time, propagation time between terminals is effectively cancelled in a relatively simple computation or electronic process to yield the time difference between clocks. Interpolation methods have been developed, however, to permit the time transfer to be made when the specified points are not contemporaneous.

A modem designed specifically for time transfer service at stations not equipped with suitable modems has been developed at NRL. This time transfer modem employs a pseudo-random code which operates at a 10 Mhz bit rate and cycles once each 810 microseconds. Ambiguities are resolved by inverting the phase of the code for one code cycle at prearranged times, such as once per minute. The NRL modem has been used experimentally via satellite between a pair of Maryland stations. Other experimental transfers were made between Maryland and Hawaii using the communications modems of the DCA stations.

The experiments have produced verified accuracies in the order of 0.1 microseconds. It is believed that an increase in accuracy will result from a planned increase in resolution of the measuring equipment. Confirmation of this increase, however, will be difficult because alternate precise time transfer methods are not generally available.

Time Transfer Via Satellite

During the past year, communications satellites have been successfully employed by the Naval Research Laboratory (NRL) for precise long-distance clock comparisons (or time transfers). It has been demonstrated by these experiments that Defense Satellite Communications System (DSCS) contains the potential for making the accurate Naval Observatory time reference economically available to many vital areas of the world and can substantially reduce the expensive transportation of calibrated atomic clocks to distant locations.

Under sponsorship of the Naval Electronics Systems Command and with the cooperation of the Observatory and the Defense Communications Agency (DCA), techniques have been produced for using certain DSCS links in a noninterfering manner, and time transfers have been made on an experimental basis. Use of DCA facilities on an operational basis is currently being implemented.

The planned time-distribution system will place Observatory-controlled, precise reference clocks at satellite communications (SATCOM) terminals in key areas (see Figure 1). Further distribution will be accomplished by short-range techniques.

A clock maintained at a SATCOM location near the Observatory will be disciplined by the Observatory standard through comparisons over a direct microwave link or by carrying clocks regularly from the Observatory to the terminal. The clock maintained at the distant SATCOM facility will be updated periodically by comparisons through the DSCS satellite. Since the clocks at the two terminals are held to a nearly constant rate by atomic frequency standards, they need not be compared continuously, but time transfers may be made daily, weekly, or monthly, depending upon the stability of the clocks, the availability of the satellite link, and the required time accuracy.

Use of the satellite system, however, is not unidirectionally beneficial. The time-reference equipment may be used to considerable advantage by the SATCOM terminals in their synchronization procedures.

Basically, the time-transfer system is not required to insert signals or disturb the operation of station equipment if the terminals are equipped with AN/URC-55 communications modems. The synchronizing signals shown being injected into the modems (see Figure 2) are used only to aid in modem synchronization and are not required for the time transfer. All signals required by the time-transfer unit are available at test points on the modem.

The basis of the time comparisons is the transmission and reception of well-defined signals that may be recognized at the transmitter and receiver. These signals are available as recognizable points in the high-speed pseudo-random code streams generated by the AN/URC-55 communications modems used at a number of the SATCOM stations. In order to avoid interference in the internal processes of the modems, no effort is made to control the code stream. The transmitted signal as generated by the modem is therefore compared with the local clock instead of being controlled by it.

The transmitter section of each modem generates a code stream (see Figure 3) and a code generator in the receiver section of the modem at the other station matches the stream as received. It is possible, then, to extract from each modem certain recognizable code generator states, or "ticks" (designated "all-one's" in the diagram), that can be treated as though they were very short pulses transmitted by one modem and received by the other. The tick at the receiver is delayed, of course, by the amount of time it takes the signal of the transmitting station to travel up to the satellite and down to the receiver at the other

station. In addition to the transmit tick and receive tick, each modem provides a "monitor" tick that corresponds to the reception of its own transmit tick after a round trip to the satellite and back.

Time-Transfer Techniques

Two methods of time transfer have been used. In the method shown in Figure 4, each station measures the time (τ_1 or τ_2) required for a round-trip to the satellite and return. (Time interval measurements are shown as doubled line segments in the timing diagram.) At the same time, station 1 compares its transmit tick with its own clock to obtain the time interval measurement t_1, and station 2 compares the corresponding receive tick with its own clock to obtain the measurement t_2. By using half the sum of the two double range measurements τ_1 and τ_2, the time of flight from station 1 to station 2 is determined and, in effect, is then added as a correction to the receive tick at station 2. The difference between the transmit reading at station 1 and the corrected receive reading at station 2 is the difference ΔC between the station 1 and station 2 clocks. After the pair of time-interval measurements, t_1 and τ_1 recorded at station 1 or the interval measurements t_2 and τ_2 recorded at station 2, are communicated to the opposite station, the clock difference ΔC may be determined by the formula shown at the bottom of the figure.

In the second time-transfer method (see Figure 5), each station measures its own transmit tick with respect to its clock and also measures the tick received from the other station with respect to the same clock. (The transmit measurements are designated by the doubled line segments I and III, while the receive measurements are designated II and IV. The time interval measurements made at station 1 are I and II, while the measurements made at station 2 are III and IV.) If the sum of measurements at one station is subtracted from the sum of the measurements taken at the other station, the result is twice the difference between the two clocks. For this method, identical equipment and procedures are employed at the two stations. Range measurements, per se, are not made, since the total range from one station to the other is involved in each station's measurements which are subtracted to obtain the clock difference, $C_2 - C_1$ given by the formula just above the timing diagram.

Time Transfer Unit

A time-transfer unit (see Figure 6) has been designed as an interface to operate automatically in the latter method and yield a single reading at each station. The difference between clocks is simply the difference between the numbers produced at the two stations.

With either time-transfer method, however, it is necessary to transmit data from one station to the other in order to make the clock-difference determination. This is normally done through the order wire between stations.

The two methods are valid, in general, only if the transmitted ticks from both stations arrive at the satellite at nearly the same time. This is ensured if the modems of the two stations have made a tick start, an easily accomplished procedure when accurate local clocks are available. Although provisions have been made in the time-transfer unit for making tick starts with the accurate clocks of the time-transfer system, under some circumstances, such a start might not be used. In this case, the ticks of the two stations might pass through the satellite at widely separated times and the motion of the satellite between the two passages could produce an unacceptable error in the time transfer.

To compensate for this error, the range measurements may be interpolated in the first method. For the second method, an interpolation procedure has also been worked out. Although it is inconvenient to use the interpolation procedures, they apparently add little to the inaccuracies of the time transfer.

Experimental Results

In February 1970, the initial time-transfer tests were run between the SATCOM stations at Brandywine, Maryland, and Ft. Dix, New Jersey, to prove the feasibility of using the pseudo-random code stream. Further tests were later run between Brandywine and the SATCOM facility at Helemano, Hawaii. Figure 7 shows a quite consistent grouping of results for that test. (The single reading approximately 0.7 μsec away from the mean could have resulted from a misread counter.) Much of the spread (approximately 0.3 μsec) may be attributed to lack of resolution in the measuring equipment, since the least significant digit of each counter was 0.1 μsec. Somewhat better results were obtained on a smaller sample later during the same test period.

In a still later series of tests a small sample of measurements yielded a standard deviation of less than 0.1 μsec. A group of measurements made by interpolation disagreed with the group made with synchronized modems by less than 0.1 μsec. During the same period, a time transfer made by physically transporting a clock to Hawaii and back to Maryland disagreed with the satellite time transfer by approximately 0.3 μsec. The clock transport, however, was not ideal because of a relatively large rate offset in the portable clock and some delay in making the comparison after its return to Maryland. In spite of these shortcomings, it appears that satellite time transfers may be relied on for an accuracy of 0.5 μsec or better. Future tests are expected to yield a more accurate picture of its performance. The measuring equipment is currently being updated to a resolution of 10 ns for these tests.

Differential time delays between transmitting and receiving terminal equipment produce small fixed time offsets. Since the individual contributions of the various station components are small and are complicated by frequency conversions, it may be difficult to analyze them individually. Instead, it will probably be necessary to make overall station comparisons to determine the absolute accuracy of the satellite time transfers. It is expected that the effects of the differential delays will prove to be less than 0.1 μsec.

Time-Transfer Modem

A modem intended specifically for time transfers between terminals not equipped with communications modems has been designed and constructed. This modem employs a pseudo-random code that operates at a 10-Mbz bit rate and has a complete code cycle length of 819 μsecs as shown in Figure 6. Since the recognizable all-one's events recur once each 819 μsecs other information must be transmitted by the modem to identify which all-one's event is the intended time tick. As seen, the transmitter responds to an initiating pulse by reversing the phase of the code throughout the next code cycle. The all-one's event that terminates that

cycle is the designated tick. At the receiver, the all-one's event that occurs at the end of the phase reversal is similarly recognized as the tick.

Time transfers are made by initiating time ticks at prearranged times, such as the beginning of each minute. The transmitted and received ticks produced by the time-transfer modem at each station are provided to the time-transfer unit, which performs the necessary time-interval measurements.

Because of the comparatively low relative velocity of DSCS satellites, it is not necessary for the signals of both stations to reach the satellite at precisely the same time. In fact, a 1-second difference in arrival time produces a time-transfer error in the order of only 0.1 μsec. Therefore, the only requirement for control of the transmissions is that they be launched within a fraction of a second of each other. When the time-transfer modem is used, this requirement is satisfied if the clocks at the two stations agree within a fraction of a second because the designated tick occurs less than 2 milliseconds after the initiating pulse.

A simplified diagram illustrating the transmission and reception processes of the time-transfer modem is shown in Figure 9. The lower section is the transmitter, which produces a pseudo-random bi-phase modulated 70-MHz signal. The code stream polarity (hence the polarity of the modulated output signal) is reversed for one code interval after the command pulse is received. The polarity inversion sequence circuit also produces the gated all-one's pulse that constitutes the transmit-time tick. Station frequency converters translate the 70-MHz signal to X-band for transmission to the satellite.

The X-band signal relayed by the satellite is converted down to 70 MHz for the modem's receiver section shown in the upper portion of the diagram. It is prevented from receiving the time tick of its own transmitter by transmitting a different code from each station. The receiver at station 1, for example, uses the station 2 transmitter code and, therefore, ignores the station 1 transmitter.

Range measurements may be made by using a common transmit and receive code at one station. The receiver then responds to the local transmitted signal after its round trip to the satellite. The time interval from the transmit tick to the receive tick then corresponds to the double-range propagation time.

In the receiver, the local conversion oscillator is bi-phase modulated with the same code that modulated the received signal. After the two codes have been aligned by a searching process, the high-speed phase inversions of the received signal are accompanied by corresponding phase inversions of the local oscillator. The intermediate-frequency output of the modulator, therefore, is a constant-phase signal except for the occasional phase-inverted code cycle that precedes a time tick.

Any misalignment of the code modulation of the received signal and the local oscillator results in an amplitude change at the intermediate frequency. By controlling the rate of the receiver code generator to maintain peak intermediate-frequency amplitude, the receiver code generator's all-1's pulses may be considered to represent accurately the all-1's pulses of the received signal.

When the phase of the received signal is inverted for a whole code cycle to designate the time tick, the phase-reversal detector of the receiver detects the change, and its filtered output gates the designated all-one's pulse from the code generator to represent the received tick. The phase-reversal detector consists of a slow-acting locked oscillator and phase detector to compare the intermediate-frequency signal with the oscillator.

The time-transfer modem (Figure 10) has been operated in the presence of interfering pseudo-random or frequency-modulated signals 20 db stronger than the desired signal. Range measurements were made at Helemano, Hawaii, under normal communications conditions with the modem power reduced approximately 15 to 20 db below the station communications channel power. It is possible, therefore, to use the modem in many cases without serious interference to normal communications.

Results of Time-Transfer Modem Experiments

Time transfers have been made between the small Brandywine TSC-54 transportable terminal and the Waldorf, Maryland, NRL satellite communications facility using the time transfer modem (see Figure 11). Although the distance between sites is an unimpressive 10 miles, the test was useful both to check out the modem and to evaluate the accuracy of time transfers in general.

The total propagation distances actually involved were as large as those for more widely separated terminals, and satellite motion and other factors were believed to be typical. The advantage of using nearby terminals is that the accuracy of the time transfer can be checked reliably.

After the time transfer was completed, the truck containing the Brandywine clock was driven to Waldorf and the clocks were compared directly. Figure 12 shows the result of the satellite comparisons. The spread is typical of transfers made at other locations.

The direct comparison made at Waldorf indicated a clock difference of 0.6 μsec. This would indicate a 0.15 μsec discrepancy with respect to the average of the satellite transfer readings. However, the disagreement may actually be less than that. The 0.6 μsec figure was based on a single reading, but a series of clock comparisons made after returning both clocks to the Laboratory about one hour later, indicated that the clocks were more nearly 0.7 μsec apart and that the discrepancy might have been less than 0.1 μs.

It can be readily concluded that the accuracy of satellite time transfer is within the ability of our present equipment to discern. An improvement in resolution obtained by using a higher counting rate in the time-transfer unit is expected to yield greater accuracy. But it will be most difficult to prove because equal or better methods are not available.

Immediate Applications of Technique

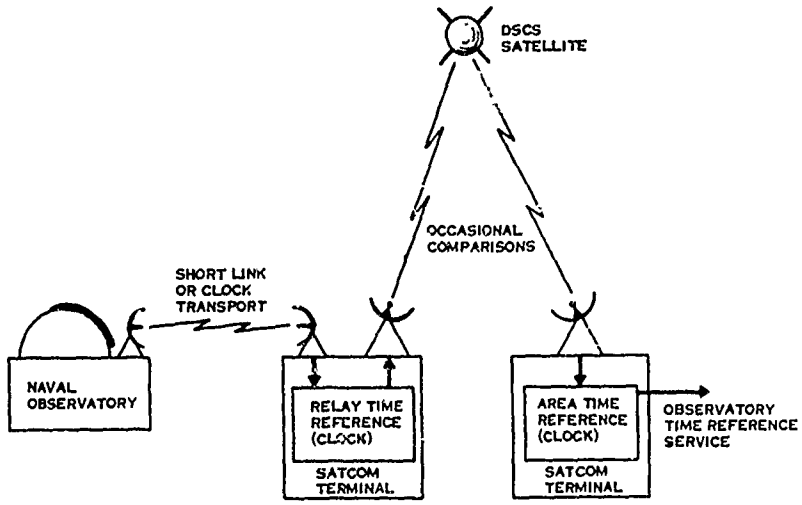
With equipment already developed it is practicable to provide Observatory time to within a small fraction of a microsecond to a number of strategically situated areas. (Figure 13).

An operational time-transfer link was established on February 15 between Brandywine, Md. and Helemano, Hawaii. A second link was completed between Hawaii and Guam shortly afterwards. This link is regularly activated by "blind" modem starts at precisely prearranged times while the Hawaii station maintains normal communications with another station.

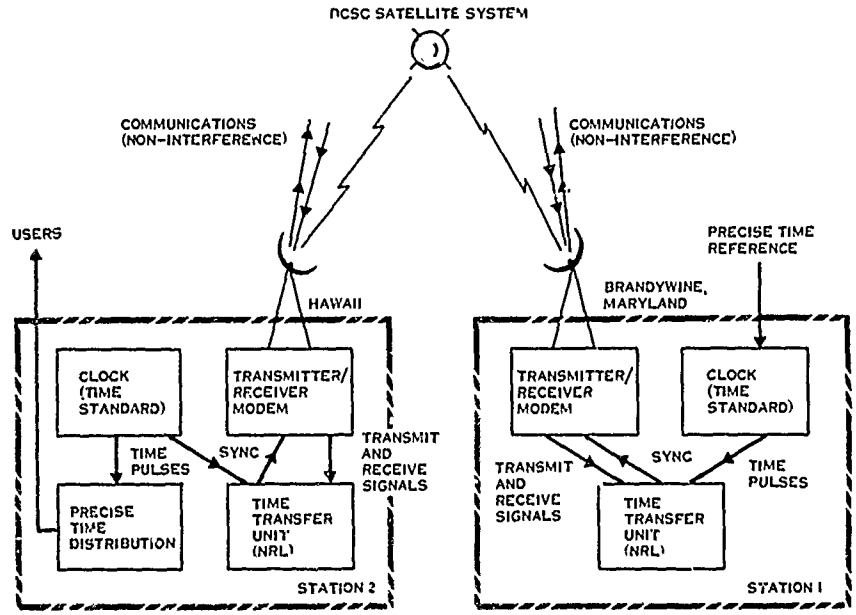
A station at Camp Roberts, California now makes regular blind-start time transfers with the Brandywine terminal while the latter is in communication with Hawaii. Within a few weeks another operational time transfer link will be established between Ft. Dix, N. J. and Landstuhl, Germany. Ft. Dix will be referenced back to the Observatory by way of blind-start comparisons with Brandywine.

These trunk-lines will soon be augmented by other links using existing communications modems where they are available and the time transfer modem where they are not.

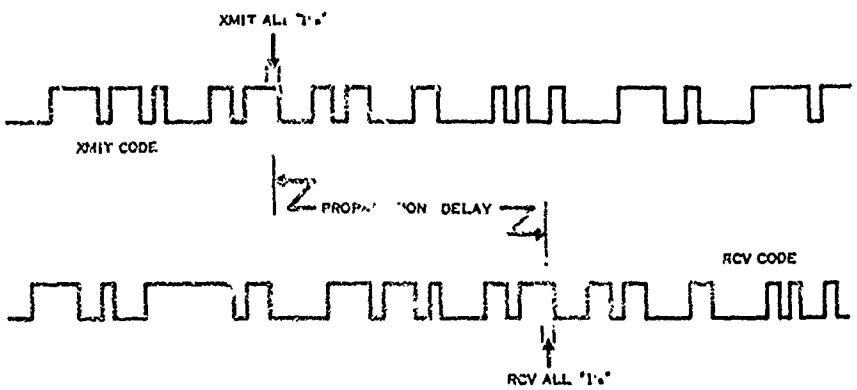
TIME REFERENCE SERVICE VIA SATELLITE

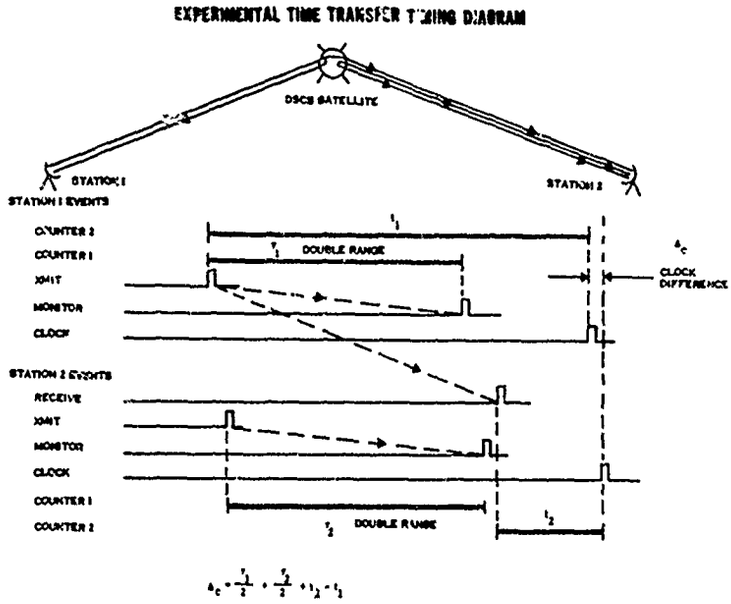


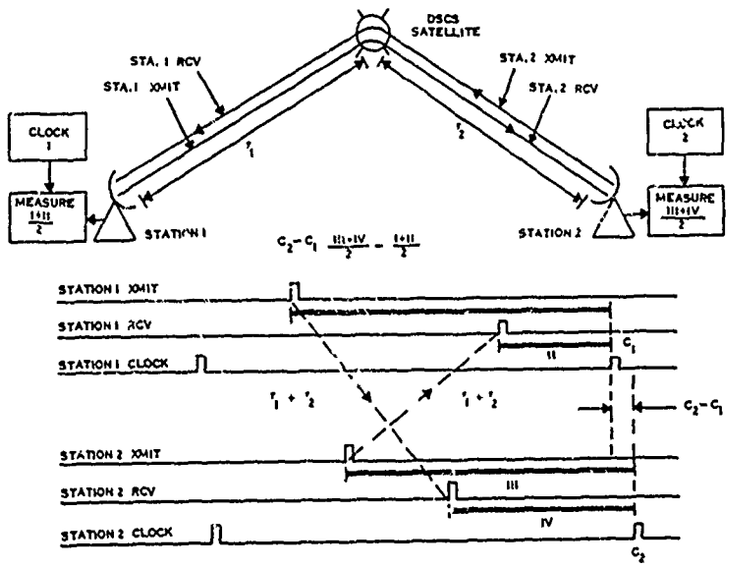
TIME TRANSFER BY COMMUNICATION SATELLITE

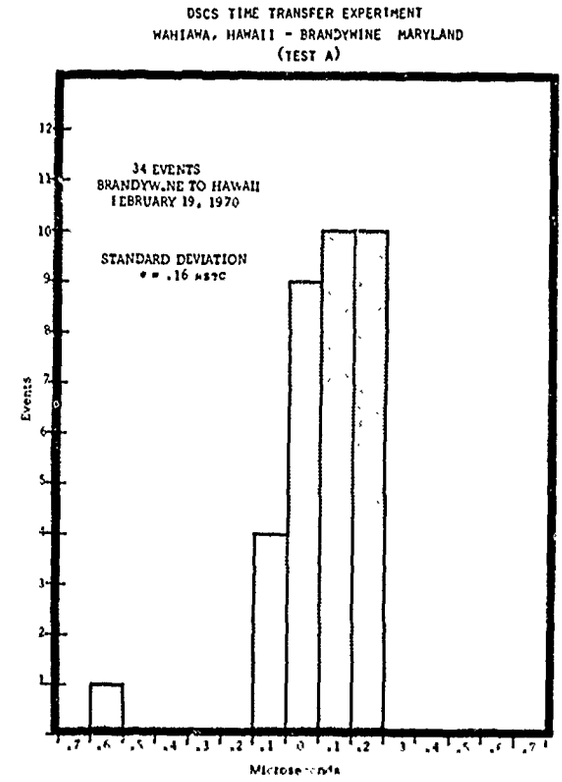


XMIT–RCV PN CODE

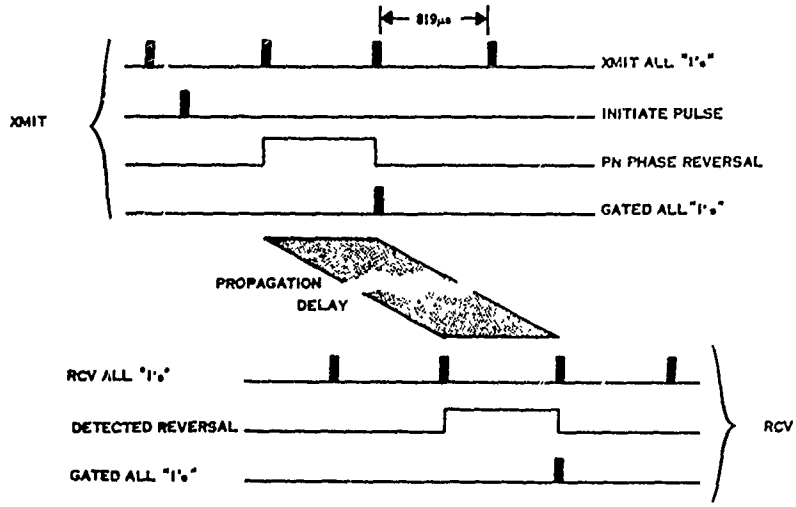



EXPERIMENTAL TIME TRANSFER TIMING DIAGRAM
DSCS SATELLITE
STATION 1
STATION 2
STATION 1 EVENTS
COUNTER 2
COUNTER 1
XMIT
MONITOR
CLOCK
DOUBLE RANGE
CLOCK DIFFERENCE
STATION 2 EVENTS
RECEIVE
XMIT
MONITOR
CLOCK
COUNTER 1
COUNTER 2
DOUBLE RANGE

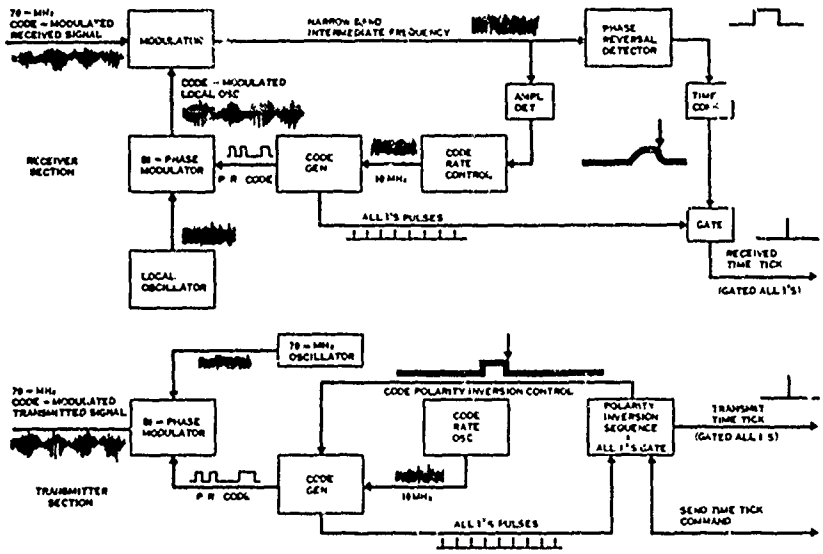

TIME TRANSFER TIMING DIAGRAM
DSCS SATELLITE
STA. 1 RCV
STA.1 XMIT
STA.2 XMIT
STA.2 RCV
CLOCK 1
CLOCK 2
MEASURE
STATION 1
STATION 2
STATION 1 XMIT
STATION 1 RCV
STATION 1 CLOCK
STATION 2 XMIT
STATION 2 RCV
STATION 2 CLOCK

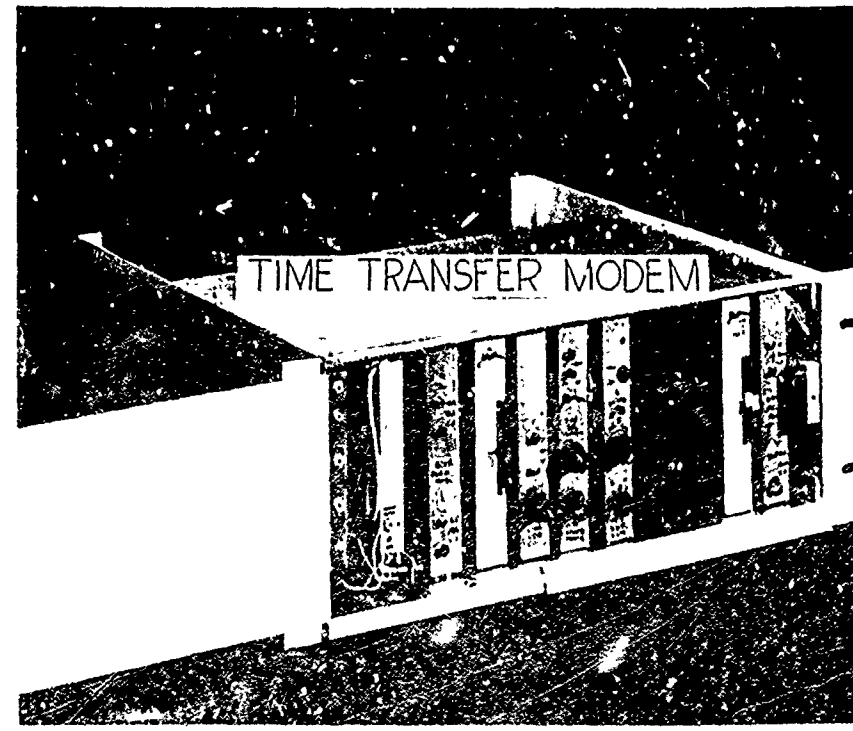

DSCS TIME TRANSFER EXPERIMENT
WAHIAWA, HAWAII - BRANDYWINE MARYLAND
(TEST A)
34 EVENTS
BRANDYWINE TO HAWAII
FEBRUARY 19, 1970
STANDARD DEVIATION
Events

SIGNAL TRANSMISSION BY TIME TRANSFER MODEM

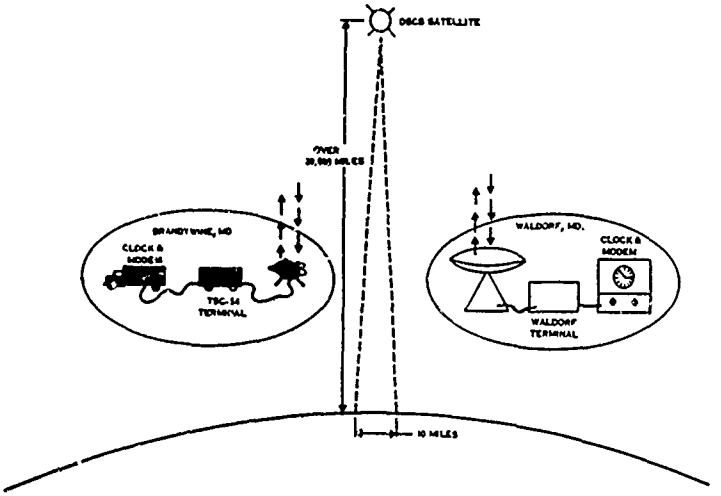


TIME TRANSFER MODEM TRANSMITTER AND RECEIVER

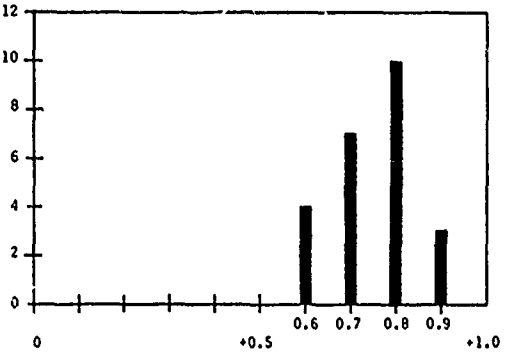




TIME TRANSFER EXPERIMENT



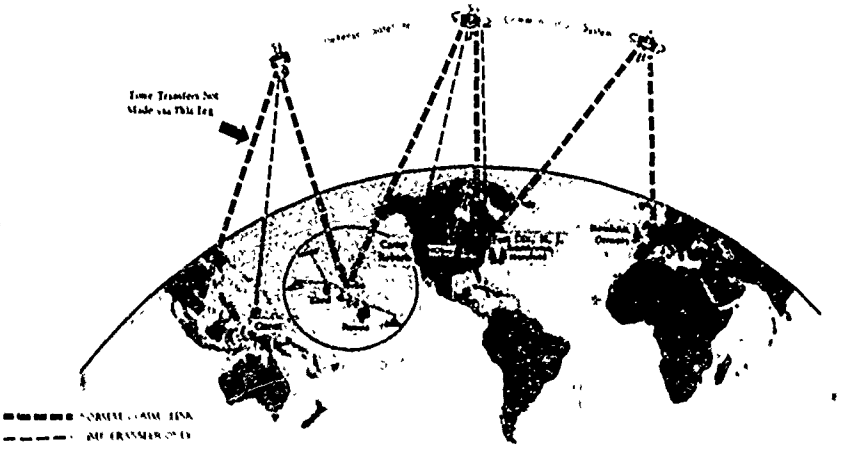
TEST RESULTS
SATELLITE TIME TRANSFER MODEM
BRANDYWINE, MD. (TSC-54 TERMINAL) TO WALDORF, MD.



Waldorf Clock - Brandywine Clock (microseconds)

28 October 1970

PRECISE TIME AND TIME INTERVAL (PTTI) – WORLD DISSEMINATION



LONG TERM ACCURACY OF TIME COMPARISONS VIA TV RADIO RELAY SYSTEMS

By

S. Leschiutta

Instituto Elettrotecnico Nazionale

This paper deals with the time comparison method that is based on TV synchronization pulses. This method was proposed by J. Tolman et al. and was successfully tested in Czechoslovakia. In more recent years experiments of a larger extent were performed in Europe and in the U.S. Furthermore, such a method is a routine operation for standard time laboratories that compare their time scales. The time comparison is between the read of counters that are located near different clocks, Each counter is started by the "second" of the local clock but is stopped by the same field synchronizing pulse of a TV transmission, as it is received at the clock location.

For long term timekeeping, the instability of the delay along the radio links is obviously a major source of errors. In 1969 a research on this subject was initiated in Italy. It is still in progress. Some of its preliminary results are given here.

Once a week two Laboratories, the Instituto Elettrotecnico Nazionale (IEN) in Turin and the Instituto Superiore e Telecomunicazioni (ISPT) in Rome, are comparing their standard clocks. They use TV signals transmitted via a microwave link, that has a length of about 742 km, with 8 relay stations.

PRECISION AND ACCURACY OF REMOTE SYNCHRONIZATION VIA PORTABLE CLOCKS, LORAN C, AND NETWORK TELEVISION BROADCASTS*

David W. Allan, B. E. Blair, D. D. Davis, and H. E. Machlan

Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80302

Summary

A comparison among three precise timing centers in the United States has been conducted for more than one year using three different synchronization methods. The timing centers involved were the United States Naval Observatory (USNO) in Washington, D. C., Newark Air Force Station (NAFS) in Newark, Ohio, and the National Bureau of Standards (NBS) in Boulder, Colorado. The three methods were cesium beam portable clocks; Loran-C transmissions from Cape Fear, North Carolina, and Dana, Indiana; and ABC, CBS, and NBC network television broadcasts common to the three timing centers.

Cesium beam portable clocks have the capability of accurately and precisely synchronizing remote clocks to within 0.1 μs. This portable clock method, which appears to be one of the most accurate currently available, is used for comparison purposes.

The Loran-C data involved a 3500 km (2180 miles) ground wave path--the longest Loran-C ground wave path that has been studied with the precision and accuracy reported herein. The long-term precision achieved via Loran-C between the three remote precise timing centers was better than 2 μs. The accuracy is limited by the 10 μs ambiguity involved in identifying the proper cycle of the 100-kHz pulse train.

The precision of maintaining remote clock synchronization using network television broadcasts was measured to be $[30 \text{ ns day}^{-1/2}]\ T^{1/2}$ over the range of T from 7 days to about 200 days, but with definite accuracy limitations caused by such factors as occasional network re-routing of the television signals.

An upper limit of the long-term frequency stabilities among the references used at the three timing centers were measured or inferred. Typically the reference at each center was composed of an ensemble of cesium beam frequency standards. The relative stabilities measured for sample times of the order of three months were a few parts in 10^{14}.

Key Words: Cesium beam standards, Frequency standards, Loran-C, Portable clocks, Time synchronization, TV timing.

*Contribution of the National Bureau of Standards, not subject to copyright.

Introduction

Time and frequency dissemination via television has received much attention during the past four years. Even though some very impressive results have already been obtained using television,[1-6] it seems still to be a pioneer field. In this paper we compare television with two other state-of-the-art methods of time and frequency dissemination--Loran-C and portable clocks--and evaluate some precision and accuracy capabilities of each.

The television method is readily available, very inexpensive, and within the majority of the continental United States it is a common source for many users. Loran-C, being well established and well known for its precision and accuracy,[7,8] has been chosen to compare with the television system. Portable clocks are used as a reference because of their precision and accuracy for remote synchronization--one of the best techniques yet available.[9] The portable clocks referred to in this paper are principally those of the United States Naval Observatory (USNO). We will exclude many other interesting areas such as VLF and satellite, as these have already been covered in some detail.[10-12]

There are three fundamental aspects of time which can be disseminated via these methods. The first is time interval which can be related to frequency--frequency being the inverse period of an oscillation. The second aspect is that of date or clock reading which has often been called epoch. We prefer the use of the word date because epoch has alternate meanings that could lead to confusion. Often we have a master clock, and we wish to communicate its date or time by some technique to a slave station located elsewhere. The third aspect is simultaneity--the practical application of which is clock synchronization, i.e., two clocks have the same reading in some frame of reference.

In principle, if we had perfect clocks, we could synchronize them once and they would remain synchronized forever. There are two basic reasons in practice why the synchronization does not persist. First of all, systematic problems such as frequency drift, frequency offset, and environmental effects on equipment often cause time dispersion. These must be analyzed and solved at each particular location. Secondly, there are different kinds of noise, or what we might call non-deterministic kinds of processes, that affect these time and frequency centers or communication systems. These processes are perhaps a little better classified and more universally present than some systematic effects. We will discuss and apply some useful statistical measures for the time and frequency dispersion of the dissemination systems in question.

Most of the data analyzed in this paper were taken by other people, and we wish to acknowledge the fine work and careful data taking and reporting of the personnel at the Bureau International de l'Heure (BIH), the Neuchâtel Observatory (ON), the U. S. Naval Observatory (USNO), and the Newark Air Force Station (NAFS), as well as of the personnel in addition to the authors at the National Bureau of Standards (NBS).

Methods of Analysis

It is often the case that data are taken at a constant repetition rate with a period of sampling, T; and each data point is an average over a time τ called the sample time. Let the total number of data points taken in a continual data set be M. Further, for every measurement system there is a high frequency cut-off usually called the measurement system bandwidth, f_B, such that noise at frequencies greater than f_B will be attenuated and non-relevant. In the past it has been common practice to compute the standard deviation as a statistical measure of such a data set:

$$\sigma_{\text{std.dev.}} = \left\{\frac{1}{M-1}\left[\sum_{i=1}^{M}(z_i - \bar{z})^2\right]\right\}^{1/2}, \qquad (1)$$

where z_i denotes the i^{th} data point and $\bar{z}$ denotes the average of all M of the z_i. For most of the noise processes that are pertinent in time and frequency, it has been shown that $\sigma_{\text{std. dev.}}$ depends upon M, T, τ and f_B,[13-15] and all of these parameters should be noted for each experiment. We have found it convenient to use the Allan variance:[13-15]

$$\sigma^2 = \langle\sigma^2(N, T, \tau, f_B)\rangle, \qquad (2)$$

where the angle brackets denote the expectation value. If N = M, then eq (2) is exactly the expectation value of the squared standard deviation.

For many pertinent noise processes, we have found that a power law spectral density is a good model, i. e ,

$$S_y(f) = \underline{A} f^{\alpha}, \qquad (3)$$

where f is the Fourier frequency and $\underline{A}$ is the intensity of the noise process. Throughout this paper y denotes the fractional frequency deviations, x denotes the time deviations, and so y is proportional to the derivative of x. Further, it has been shown that if N, f_B, and the ratio $r = T/\tau$ are held constant, then σ_y^2 is equal to $\underline{a}\tau^{\mu}$ with

$$\mu = \begin{cases} -\alpha - 1 & -3 < \alpha \leq 1 \\ -2 & \alpha > 1 \end{cases}, \qquad (4)$$

and with $|\tau f_B|^{|\alpha-1|} \gg 1$.[10,13] For a more detailed explanation of the N, T, and τ dependence, the relationship between μ and α, and the relationship between the frequency domain and time domain coefficients, $\underline{A}$ and $\underline{a}$, see Refs. 14 and 15.

A very useful time domain measure of frequency stability, which has been recommended by the IEEE subcommittee on frequency stability,[14] is defined as follows:

$$\sigma_y(\tau) = \langle\sigma_y^2(N = 2, T = \tau, \tau, f_B)\rangle^{1/2}. \qquad (5)$$

This measure has some very convenient experimental and theoretical characteristics. For example, with M values of $\bar{y}_i$ an estimate of $\sigma_y(\tau)$ is:

$$\sigma_y(\tau) = \left[\frac{1}{2(M-1)}\sum_{i=1}^{M-1}(\bar{y}_{i+1} - \bar{y}_i)^2\right]^{1/2}, \qquad (6)$$

where

$$\bar{y}_i = \frac{x_{i+1} - x_i}{\tau} \qquad (7)$$

and the interval between each discrete time measurement x_i is τ. Also once $\sigma_y(\tau)$ has been calculated, the time dispersion may be estimated by simply calculating the product $\sigma_x(\tau) \approx \tau \cdot \sigma_y(\tau)$.[14] This estimate is good for white noise FM but is approximately a factor of 1.3 too optimistic for flicker noise FM and is approximately a factor of 1.7 too pessimistic for flicker noise PM with respect to an optimum prediction routine.[16]

An operational measure of time dispersion is $\langle\sigma_x^2(2, T, \tau, f_B)\rangle$. This has the nice physical interpretation that it is a measure of the time dispersion at a time T after synchronization. The sample time τ in this case is usually constrained by the measurement system to be $1/(2\pi f_B)$, and this measure has a smaller range of convergence, i. e. , $\alpha > -1$ unless properly normalized (see bias function B_2 in Ref. 15)

The above statistical measures will be used to calculate the precision for the data used in this paper From these analyses we will draw some conclusions regarding the relative precision of the three time and frequency dissemination techniques herein discussed, i. e. , network television, Loran-C, and portable clocks Accuracy, on the other hand, may be defined as follows for frequency it is the confidence with which a frequency is known with respect to the currently defined resonance in cesium 133; and for time it is the confidence with which a date is known with respect to a reference time scale such as UTC(BIH).

Time and Frequency Stability of Network TV

Description of TV Line-10 Timing System

After the work of Tolman and others,[1-5] the TV line-10 system was developed by NBS as a passive means of comparing precision clocks, remotely located but periodically tied to common broadcasts from an originating network.[6] An overview of the system is shown in Fig. 1. These broadcasts originate from the New York City studios of any or all of the three commercial TV networks (ABC, CBS, and NBC). The originating networks incorporate independent atomic frequency standards (rubidium) to stabilize their transmissions. The New York signals, broadcast without auxiliary time coding, traverse varied and long paths at microwave frequencies. This relay system is a chain of broadband radio links encompassing the continental United States at line-of-sight distances of some 40 to 60 km between repeaters. The microwave relay system carrying over 95% of the U. S. intercity television programs is known as the TD-2 system.[17] At a terminating station, such as an affiliate local transmitter, the microwave signal from the applicable repeater station is converted to a video signal and re-transmitted by VHF or UHF (commercial TV) to a local service area. Reception points of such broadcasts for our data were the U. S. Naval Observatory (USNO), Washington, D. C. · Newark Air Force Station (NAFS), Newark, Ohio; and the National Bureau of Standards (NBS), Boulder, Colorado.

This version of TV timing uses line-10 of the odd field in the 525-line system M as a passive transfer pulse.[3,6] This pulse occurs during the blanking retrace interval between successive fields; line-10 was chosen for timing as it is the first horizontal sync pulse following the equalizing and vertical sync pulses and therefore is easy to identify with simple logic circuits. Figure 2 shows a typical equipment alignment for line-10 synchronization. Almost any type of television receiver, black and white or color, is suitable for reception of the signals used for timing. The auxiliary equipment required can cover a broad range of specifications, but typically would include a line-10 synchronized pulse generator (available at a cost of about $165), a digital counter-printer (about 0.1 μs resolution), and a clock whose frequency is known to an accuracy of a few parts in 10^9 and which has an output of 1 pulse per second (1 pps). A functional block diagram of the line-10 identification circuitry is shown in Fig. 3.

The timing system employed to collect the data used in this paper works as follows: At the same date, to an accuracy which is much better than the 17 ms needed, counters are started at all laboratories with a 1 pps tick from their local reference atomic clocks. Near this time a line-10 horizontal sync pulse is broadcast from one of the New York City originating TV transmitters. After diverse delays through both common and separate microwave links, the sync pulse--received by the laboratories at different times due to the delay and clock differences--stops the corresponding counters. The difference, then between each pair of counter readings at any two receiving laboratories, remains constant except for instabilities in the propagation delay and/or instabilities or relative frequency offsets among the reference atomic clocks.

Similarly, any laboratory can compare their clocks to the UTC(USNO) and UTC(NBS) scales through a duplicate reception-recording system once their clock has been set to within 17 ms of either of the above, and the propagation path has been calibrated. (As the period of one modular TV frame is about 33 ms, it is necessary to resolve ambiguity at the receiving site to about 17 ms.) NBS distributes the line-10 daily measurements in terms of UTC(NBS) in the Monthly NBS Time Service Bulletin.[18] The USNO distributes line-10 data in terms of UTC(USNO-MC) in the weekly Series 4 Time Services Bulletin.[19] These publications will publish future changes, modifications to the TV system, or other factors affecting a user in the field.

Advantages of line-10 timing include 1) simplicity and minimum cost of comparison equipment; 2) low cost of maintaining synchronization with long range precision of better than 10 μs; 3) no effect on regular TV networks and with no external cost to user; 4) three TV networks with atomic clock references provide redundancy and backup data in case one TV channel shows a microwave reroute; and 5) a method for simultaneous maintenance of μs synchronization of several clocks diversely located within the service area of a common transmitter, without regard to national programming. These advantages must be tempered by such factors as 1) microwave paths can be interrupted without notice; 2) there is limited simultaneous viewing time of nationwide network programs; 3) present network distribution does not allow common programming with west coast transmission lines, although local synchronization from a common transmitter can be effected; 4) system is not compatible with tape delayed programs, and 5) the system ambiguity is 33 ms. (Note that time-of-day is "ambiguous" to one day.)

Stability of the USNO, NBS TV Path

The TV paths being considered are from two to four thousand kilometers in length. The particular problem mentioned above of an occasional TV network re-route will cause an effective change in the delay. So it is highly advantageous to use all three networks so that such a change can readily be identified. Conveniently, we do have three networks so that outliers and delay changes are easily recognized. During the analysis period studied, which was from 25 June 1969 to 30 December 1970, there were only about two network delay changes per year per network, so it is not a serious inconvenience. The television time measurements using the line-10 method were made at 2025 UT, 2026UT, and 2027 UT (one hour earlier during Daylight Savings Time) on NBC, CBS, and ABC, respectively and nominally every work day at each of the three laboratories.

Figure 4 is a plot of the fractional frequency stability, $\sigma_y(\tau)$ versus the sample time τ in days for the TV paths between Washington, D. C. and Boulder, Colorado, for each of the three networks (this assumes the reference time scales are negligible). An ensemble of commercial cesium beam frequency standards and dividers to generate atomic time (AT) were used at each location as the 1 pps time reference, i.e., AT(USNO-MEAN) and AT(NBS).[20-23] The dashed line in Fig. 4 corresponds to flicker noise phase modulation and appears to be a reasonable model for the fluctuations

between Washington, D. C., and Boulder, Colorado. Calculating the time dispersion, $\sigma_x(\tau)$ for the dotted line gives the following equation:[14]

$$\sigma_x(\tau) = 62 \text{ ns} [6 + 3 \log_e(2\pi \tau f_B) - \log_e 2]^{1/2}, \quad (8)$$

where τ is in seconds; f_B for a color TV is about 3.2 MHz.

The fractional frequency difference, y, between AT(USNO-MEAN) and AT(NBS) calculated over the period of analysis was:

$$\frac{\nu_{AT(USNO\text{-}MEAN)} - \nu_{AT(NBS)}}{\nu_{AT(NBS)}} = \begin{cases} 4.48 \times 10^{-13} & \text{via ABC} \\ 4.43 \times 10^{-13} & \text{via CBS} \\ 4.56 \times 10^{-13} & \text{via NBC.} \end{cases} \quad (9)$$

The precision of these measurements is about $\pm 3 \times 10^{-14}$ as may be seen from the stability measured at τ = 224 days.

Since there are three essentially independent networks, they can be combined optimally by weighting each one inversely proportional to its variance. We chose the variance at τ = 7 days, since it has the best confidence, to calculate the weights of 0.52, 0.38, and 0.10 for ABC, CBS, and NBC, respectively. This is optimum in the sense of giving a minimum variance, and it should be noted that these coefficients need to be calculated for each pair of receiving laboratories or effectively for each TV path. The squares in Fig. 4 represent the stability using this optimum weighting procedure. The squares in Fig. 5 show the time fluctuations of the weighted three network TV data between USNO and NBS as measured each Wednesday. Each of these points represents a 40s average. A one second measurement gives almost as good precision, but the averaging allows one to recognize outliers. The interval between the measured line-10 horizontal sync pulses is 1.001s and hence there is a walk between a standard 1 pps and the line-10 pulse of 1 ms per second (modulo 33 ms). This is easily accounted for in the counter printer system shown in Fig. 2.

TV and Cesium Beam Stability at NAFS

The same line-10 TV network method was employed as outlined above at Newark Air Force Station (NAFS) in Newark, Ohio. Their time reference was a commercial cesium beam frequency standard and clock. Both USNO and NBS data were used to study the path stability between Washington, D. C., and Newark, Ohio, and between Boulder, Colorado, and Newark, Ohio, over the periods from 24 September 1969 to 16 December 1970 and from 17 September 1969 to 30 December 1970 respectively.

Figure 6 shows the fractional frequency stability $\sigma_y(\tau)$ for both paths using all three TV networks optimally weighted. The squares are the stability for the USNO, NAFS TV path with weights of 0.36, 0.57, 0.07 for ABC, CBS, and NBC respectively. The circles are the stability for the NBS, NAFS TV path with weights of 0.29, 0.36, and 0.35 for ABC, CBS, and NBC respectively.

The slope indicated by $\alpha = 0$ is probably the noise of the TV line-10 time transfer system, but the type of noise is unexpected, i.e., white frequency FM or random walk of phase noise. Calculating $\sigma_x(\tau) = \tau\sigma_y(\tau)$ gives:

$$\sigma_x(\tau) = (0.13 \ \mu\text{s day}^{-1/2})\, \tau^{1/2} \quad (10)$$

where τ is in days. An explanation of this random walk of phase noise could be some step changes in the delay for which there was no accounting.

Note that the stability gets worse for τ larger than 100 days with a maximum at τ equal to about 1/2 year for both the squares and the circles. This part of the stability plot is probably due to a seasonal or annual effect on the time reference cesium standard at NAFS.

Wiener Filtering of TV Data

In Fig. 6 the dashed line representative of $\alpha = 0$ (white noise FM) appears to be a good model for the instabilities in the TV data over the two paths mentioned above. If one can assume that the dashed line representative of $\alpha = -2$ (random walk of frequency noise) is a good noise model for the cesium beam reference standard, then it has been shown that a Wiener filter may be applied to the data.[11, 24] Assuming that the $\alpha = 0$ process is noise and the $\alpha = -2$ process is signal, i.e., we wish to have a best estimate of the behavior of the cesium reference as observed through the noise of the TV line-10 time transfer system, then the filter takes on a very simple form, i.e., an exponential. We mean by best estimate a minimum mean squared error for the signal $\langle[\hat{s}(t) - s(t)]^2\rangle$ where $s(t)$ is the true behavior of the cesium beam reference standard and $\hat{s}(t)$ is the Wiener filtered estimate.

The models for the noise and signal are.

$$S_{y(noise)}(f) = A\, f^0, \quad (11)$$

and

$$S_{y(signal)}(f) = B\, f^{-2} \quad (12)$$

respectively, and the form of the Wiener futer for the discrete case is as follows:

$$\hat{x}_i = \frac{\sum_{j=-\infty}^{i} x_j\, e^{-\frac{i-j}{\xi}}}{\sum_{j=-\infty}^{i} e^{-\frac{i-j}{\xi}}}, \quad (13)$$

where

$$\xi = \frac{1}{2\pi}\sqrt{\frac{A}{B}} \tag{14}$$

and ξ is normalized to the same units as i and j, i. e., days, weeks, etc. A convenient recursive filter that approximates eq (13) is:

$$\hat{x}_i = \frac{1}{\xi+1}[x_i + \xi\,\hat{x}_{i-1}]. \tag{15}$$

For the signal and noise processes given by eqs 11 and 12, using Ref. 14, ξ takes on the value

$$\xi = \frac{1}{\sqrt{3}}\,\tau_I\,, \tag{16}$$

where τ_I is the value of τ corresponding to the intercept of the dashed lines in Fig. 6. For these particular data sets, ξ had values of 7 weeks and 9-1/4 weeks for the Washington, D. C. to Newark, Ohio, and the Boulder, Colorado to Newark, Ohio, paths respectively.

The dots in Fig. 7 are the weighted three network values measured each Wednesday for each of the above two paths. The solid lines are the result of an application on these data of the Wiener filter given by eq 13 with the sum being taken over about 1-1/2 time constants, i. e., the past 10 and 14 values respectively for the above two paths. Note the slope of one part in 10^{13} and the strong correlation between the two paths both before and after filtering.

Time Dispersion of Optimum Processed TV Data

Taking the difference between the two solid lines in Fig. 7 gives us a filtered estimate of time fluctuations between AT(USNO-MEAN) and AT(NBS). If we now apply a Wiener filter to the TV line-10 data plotted in Fig. 5, we have a direct path filtered estimate of the same fluctuations. Taking the difference between these two estimates leaves as a residual the filtered noise of the TV line-10 time transfer system and the associated measurement equipment. This residual is plotted in Fig. 8. Note, the data fall within a very sensitive vertical range of ± 1 μs. The Wiener filter applied to the USNO, NBS path[11] assumes that the instabilities of the TV line-10 time transfer system are characterized by flicker noise PM($\alpha = 1$, see Fig. 4), and that the signal, AT(USNO-MEAN) - AT(NBS), is limited by flicker noise FM($\alpha = -1$) at a level for $\sigma_y(\tau)$ of about 1×10^{-13}. The latter assumption, though inconsistent with the 3×10^{-14} value of $\sigma_y(\tau)$ from Fig. 4, was based on two reasons: first, frequency changes of this order have been observed between the two scales (see Fig. 5, November, 1970); and second, the filter has a similar time constant to the other two paths.

When the data in Fig. 8 are analyzed using the time dispersion measure previously discussed, we get the results shown in Fig. 9. The white noise FM model represented by the dashed line seems to be a reasonable model at a level of:

$$\sigma_x(2, T, \tau, f_B) = (30\,\text{ns day}^{-1/2})T^{1/2}, \tag{17}$$

with T having a range from 7 days to about 200 days.

For comparison we analyzed the same data to determine $\sigma_x(\tau) \equiv \tau\,\sigma_y(\tau)$:

$$\sigma_x(\tau) = (46\,\text{ns day}^{-1/2})\tau^{-1/2}, \tag{18}$$

where τ is in days. Using Refs. 14 and 15, the theoretical ratio of $\sigma_x(\tau)$ to $\sigma_x(2, T, \tau, f_B)$ may be calculated for white noise FM as equal to $\sqrt{2} = 1.414$, whereas the experimental value from eqs (17) and (18) is 1.5. For either equation the dispersion is less than 1 μs after one year.

Time and Frequency Stability of Loran-C

Stability of USNO, NBS Loran-C Path

The circles in Fig. 5 represent the time difference fluctuations AT(NBS) - AT(USNO-MEAN) via a 3,500 km Loran-C path going from Cape Fear, North Carolina, to USNO and to Dana, Indiana, and from Dana, Indiana, to Boulder, Colorado. Dana, Indiana is phase controlled to within 0.2 μs with respect to Cape Fear using as the phase reference point Warner Robins AFB in Georgia.[26] The circles represent the Loran-C measurements as made every fifth day. The triangles are the USNO portable clock trips between USNO and NBS with a reported accuracy of date transferral of 0.1 μs.[19]

The circles in Fig. 10 are the fractional frequency stability, $\sigma_y(\tau)$ for the Loran-C data plotted in Fig. 5. The circles in Fig. 11 are an estimate of the time dispersion, $\sigma_x(\tau) \equiv \tau\,\sigma_y(\tau)$ for the same data. The squares in Fig. 11 are the same estimate of time dispersion for the TV line-10 data shown in Fig. 5.

Stability of European and Atlantic Loran-C Paths

Some data were made available at the courtesy of the Neuchâtel Observatory (ON) using Loran-C to compare two commercial cesium standards located at ON each against UTC(BIH) and UTC(USNO). Given a set of measurements between three independent standards, the stability of each may be estimated.

Using this approach, while realizing that part of the Loran-C path is common to the two ON cesium standards, an estimate of $\sigma_y(\tau)$ was made for the USNO, ON path across the North Atlantic and to Neuchâtel, Switzerland, represented by the pluses (+) in Fig. 10. Similarly an estimate of $\sigma_y(\tau)$ was made for the BIH, ON Loran-C path and are represented by the triangles (Δ) in Fig. 10. It is interesting that the much longer Atlantic Loran-C path from USNO to ON is more stable than the USNO, NBS continental Loran-C path.

Loran-C and TV Stability via BIH

The BIH maintains the International Atomic Time Scale IAT(BIH). This scale is based on a weighted set of the atomic time scales of the major laboratories of the world which keep such scales and which have a date communication link to the BIH via Loran-C or TV.

Using the data published by the BIH in Circular-D, we performed a stability analysis of the time scale data for these major laboratories versus the BIH. The four most stable time reference laboratories were chosen in order to see the instabilities in the Loran-C system and the results are plotted in Fig. 12. It should be noted that IAT(BIH) is dependent upon each of the contributing time scales, hence the stabilities shown are nominally optimistically biased.

The atomic time scale of the Physikalisch-Technische Bundesanstalt (PTB) is communicated to the BIH via European TV and Loran-C from Braunschweig, Federal Republic of Germany, to Paris, France. The Royal Greenwich Observatory (RGO) is in Herstmonceux, England.

It is interesting to note the comparable stabilities of the four laboratories even though they are located at greatly different distances over the earth. The white noise FM is typical of cesium beam frequency standards but for much smaller τ values and at much lower levels. This noise is apparently nominally representative of the Loran-C and TV instabilities as observed at the BIH except for the NBS data which is more nearly represented by flicker noise PM. The white noise FM again implies step changes in delay for which there is no accounting.

Time Accuracy of TV Line-10, Loran-C, and Cesium Portable Clocks

The ordinate shown in Fig. 5 for the TV line-10 system, for the Loran-C, and for the portable clocks is arbitrary since over these paths only the portable clocks have accurate date transferral capabilities. For the paths being considered both the TV line-10 system and Loran-C need a path delay calibration in order to maintain sub-microsecond synchronization. Loran-C delays can often be calculated to better than 1 μs for areas within ground-wave coverage. For the TV line-10 system the path delay may be readily calculated to within about 1 μs only when both receiving points are line-of-sight to the same TV transmitter. On the other hand, the cycle ambiguity for Loran-C can be 10 μs whereas it is only 33 ms for the TV line-10 system. This means that the cycle ambiguity could be removed very easily on the TV line-10 system using WWV whereas the 10 μs ambiguity of Loran-C can be more difficult to remove.

Figure 11 shows the time dispersion characteristics for cesium beam portable clocks presently available. Note that the dispersion degrades to be that of either TV or Loran-C for τ values of a few weeks. For portable clocks τ may be interpreted as the time since the last date calibration. It is obvious for this and other reasons that portable clock trips should be made quickly for the best accuracy in date communication.

Short-Term Frequency Stability of TV, Loran-C, and Portable Clocks

It is often desirable to have available a reference standard frequency for calibrating the frequency of a clock, or of a frequency counter, etc. In color television broadcasting, a color "subcarrier" of 63/88·5 MHz (3.58 ... MHz) is transmitted on the VHF or UHF signal. It is used as a reference signal in the color television receiver to demodulate the chrominance sidebands. Since the major U. S. networks generate the color subcarrier with rubidium frequency standards, this color subcarrier may be used as a reference standard frequency. Frequency stability measurements of the color subcarriers of all three major U. S. networks (originating in New York) have been made at the NBS laboratories in Boulder.[24] In most cases it appears that the overall measurements system is capable of resolving the 3.58 ... MHz time (phase) differences to less than 10 nanoseconds. This corresponds to determination of frequency difference of about one part in 10^{11} in 17 minutes. NBS designed instrumentation both to synthesize the output of a 1 or 5 MHz local frequency standard to 3.58 ... MHz and to compare phases of the local synthesized signals to the received subcarrier frequency. A plot of the stability of some of the best data received in Boulder, Colorado, are represented by the squares marked CBS in Fig. 13. Typically the stability was a factor of two or three times worse than this. The Dana, Indiana Loran-C stability as monitored in Boulder, Colorado, is represented by the circles, and the portable clock stability by the triangles.

It is interesting to compare the relative stabilities (precision) of the three methods for a sample time, τ, of about 200 s. The values of $\sigma_y(\tau)$ are about 10^{-10}, 10^{-11}, 4×10^{-12} for Loran-C TV color subcarrier, and cesium portable frequency standard, respectively. The TV color subcarrier provides a very inexpensive and precise method for frequency calibration. The accuracy of a measurement in all three cases is limited by the accuracy of the reference standard employed as well as by the precision of the measurement. To improve the usefulness of the TV color subcarrier method it would be an easy matter for NBS to publish daily measurements of the absolute frequencies of the three networks' rubidium gas cells.

Figure 13 also shows for comparison purposes the previous data discussed comparing AT(USNO-MEAN) and AT(NBS) via Loran-C and via TV line-10 time transfer system The circles at the right are the stabilities via Loran-C and the squares at the right are the stabilities via line-10 TV.

Conclusions

The three network TV line-10 systems properly filtered may be used in a large majority of the United States to keep clocks synchronized to within an rms precision of:

$$\sigma_x(\tau) = (46 \text{ ns day}^{-1/2})\, \tau^{-1/2} \qquad (19)$$

with τ at least in the range from 7 days to about 200 days. The clocks are assumed to have been synchronized previously. The TV color subcarrier may be used as a frequency reference with a precision capability of:

$$\sigma_y(\tau) = (3.5 \times 10^{-10}\ s^{2/3})\ \tau^{-2/3} \qquad (20)$$

where τ has at least the range of values from $10s \leq \tau \leq 200s$.

The long-term fractional frequency stability of Loran-C and the three network TV line-10 systems are comparable with the Loran-C stability at a level of about:

$$\sigma_y(\tau) = (1.9 \times 10^{-12}\ day^{2/3})\ \tau^{-2/3} \qquad (21)$$

where τ has at least the range from 1 day $\leq \tau \leq$ 200 days Both systems provided precision capabilities of a few parts in 10^{14} for sample times of one-half year and longer and with rms time dispersions of less than 1 μs after one year.

Acknowledgements

We wish to thank Dr. James A. Barnes, Dr. Donald Halford, and Mr. Gunther Kramer for the helpful criticisms and comments of this paper. The assistance of Glynetta Perrymore, Dr. Bernie Wieder, and James Barnaba in the data-taking and processing has been invaluable. We also wish to acknowledge the fine work and speedy processing of this manuscript by our secretaries, Mrs. Eddyce Helfrich and Mrs. Donna Stolt.

References

1. J. Tolman, V. Ptáček, A. Souček, R. Stecher, "Microsecond Clock Comparisons by Means of TV Synchronizing Pulses," IEEE Trans on Instrumentation and Measurement, Vol. IM-16, No. 3, 1967, pp. 247-254.

2. D. W. Allan, S. Leschiutta, and G. Rovera, "TV Frame Pulses Used for Precision Time Synchronization and Their Noise Distribution," Alta Frequenza (Letter), Vol. XXXIX, No. 5, May 1970, p. 482 (p. 180 English).

3. D. D. Davis, J. L. Jespersen, and G. Kamas, "The Use of Television Signals for Time and Frequency Dissemination," Proc. IEEE (Letters), Vol. 58, No. 6, June 1970, pp. 931-933.

4. P. Parcelier, "Time Synchronization by Television," IEEE Trans. on Instrumentation and Measurement, Vol. IM-19, No. 4, November 1970, pp. 233-238.

5. S. Leschiutta, "Long Term Accuracy of Time Comparisons via TV Radio Relay Links," (Abstract) Program of the 25th Annual Frequency Control Symposium, Fort Monmouth, N.J., April 1971, p. 39.

6. D. D. Davis, B. E. Blair, and J. F. Barnaba, "Long-Term Continental U. S. Timing System via Television Networks," IEEE Spectrum (in press).

7. P. E. Pakos, "Use of the Loran-C System for Time and Frequency Dissemination," Proc. of the 23rd Annual Symposium on Frequency Control, Ft. Monmouth, N.J., 6-8 May 1969, pp. 236-247.

8. L. D. Shapiro, "Time Synchronization from Loran-C," IEEE Spectrum, Vol. 5, No. 8, August 1968, pp. 46-55.

9. L. N. Bodily and L. C. Hyatt, "'Flying Clock' Comparisons Extended to East Europe, Africa, and Australia," Hewlett-Packard J., Vol. 19, No. 4, December 1967, pp. 12-20.

10. D. W. Allan and J. A. Barnes, "Some Statistical Properties of LF and VLF Propagation," AGARD Conf. Proc. No. 33, Phase and Frequency Instabilities in Electromagnetic Wave Propagation, (Proc. AGARD/EPC 13th Symp., Ankara, Turkey, Oct. 9-12, 1967), K. Davies, Ed., Chapter 15 (Technivision Services, Slough, England), July 1970, pp. 219-230.

11. Alain Guétrot, Lynne S. Higbie, Jean Lavanceau, and David W Allan, "An Application of Statistical Smoothing Techniques on VLF Signals for Comparison of Time Between USNO and NBS," (Summary) Proc. of the 23rd Annual Symposium on Frequency Control, Fort Monmouth, N.J., May 1969, p. 248. Also unpublished thesis. University of Colorado (1967).

12. J. L. Jespersen, L. E. Gatterer, D. W. Hanson, and W. F. Hamilton, "Artificial Satellites as a Means of Time Dissemination," Proc. International Conf. on Space and Communication, Paris, France, 29 March - 2 April 1971, 9 pages (in press).

13. D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. IEEE, Vol. 54, No. 2, February 1966, pp. 221-230.

14. J. A. Barnes et al., "Characterization of Frequency Stability," NBS Technical Note 394, October 1970; also published in IEEE Trans. on Instrumentation and Measurement, Vol. IM-20, No. 2, May 1971, pp. 105-120.

15. J. A. Barnes, "Tables of Bias Functions, B_1 and B_2, for Variances Based on Finite Samples of Processes with Power Law Spectral Densities," NBS Technical Note 375, January 1969.

16. J. A. Barnes and D. W. Allan, "An Approach to the Prediction of Coordinated Universal Time," Frequency, Vol. 5, No. 6, November/December 1967, pp. 15-20.

17. U. S. Berger, "TD-2: Eighteen Years and Still Growing," Bell Lab. Record, Vol. 46, July/August 1968, pp. 210-216.

18. NBS, "Time and Frequency Services Bulletin," issued monthly to users on basis of need. Inquiries may be addressed to Frequency-Time Broadcast Services Section, Time and Frequency Division, NBS, Boulder, CO 80302.

19. USNO, "Daily Phase Values - Series 4," issued weekly to users on basis of need. Inquiries may be addressed to Director, Time and Service Division, U. S. Naval Observatory, Washington, DC 20390.

20. John B. Milton, "Standard Time and Frequency: Its Generation, Control, and Dissemination from the National Bureau of Standards Time and Frequency Division," NBS Technical Note 379, August 1969.

21. David W. Allan, James E. Gray, and Howard E. Machlan, "The NBS Atomic Time Scale System: AT(NBS), SAT(NBS), and UTC(NBS)," (Summary) Proc. of the 24th Annual Symposium on Frequency Control, Ft. Monmouth, N.J., 27-29 April 1970, p. 361.

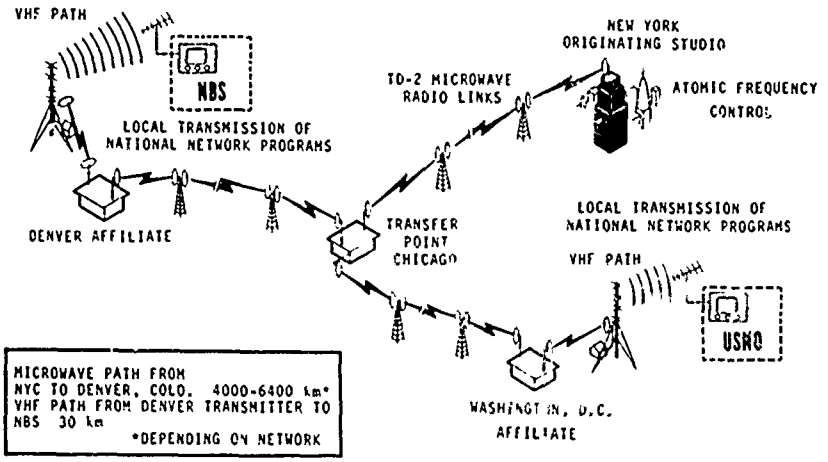
22. G.M.R. Winkler, R. G. Hall, and D. B. Percival, "The U. S. Naval Observatory Clock Time Reference and the Performance of a Sample of Atomic Clocks," Metrologia, Vol. 6, No. 4, October 1970, pp. 126-134.

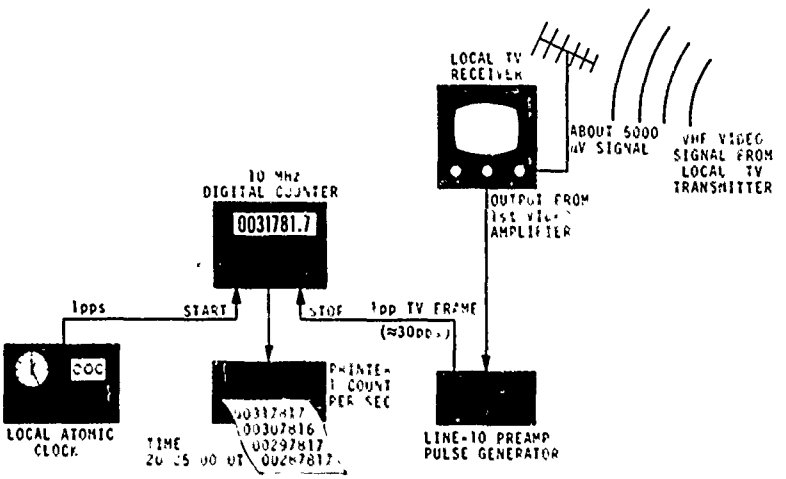
23. David W. Allan and James E. Gray, "Comments on the October 1970 Metrologia Paper, 'The U.S. Naval Observatory Clock Time Reference and the Performance of a Sample of Atomic Clocks,'" Metrologia (Letter to the Editor), April 1971.

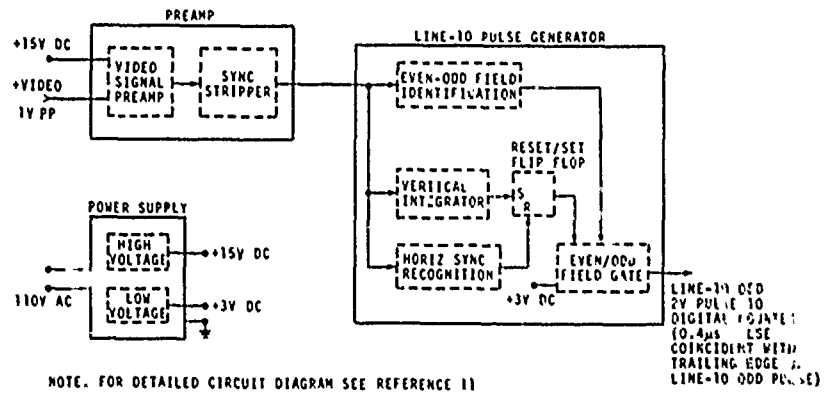
24. Y. W. Lee, Statistical Theory of Communication, John Wiley and Sons, New York, Chapter 14, 1960, p. 383.

25. D. D. Davis, "Using the TV Color Subcarrier as a Stable Frequency Source," Electronics, Vol. 44, No. 9, 10 May 1971.

26. Robert Doherty, Private Communication.







NOTE. FOR DETAILED CIRCUIT DIAGRAM SEE REFERENCE 11

Fig. 3 - Functional block diagram of TV line-10 identification circuitry

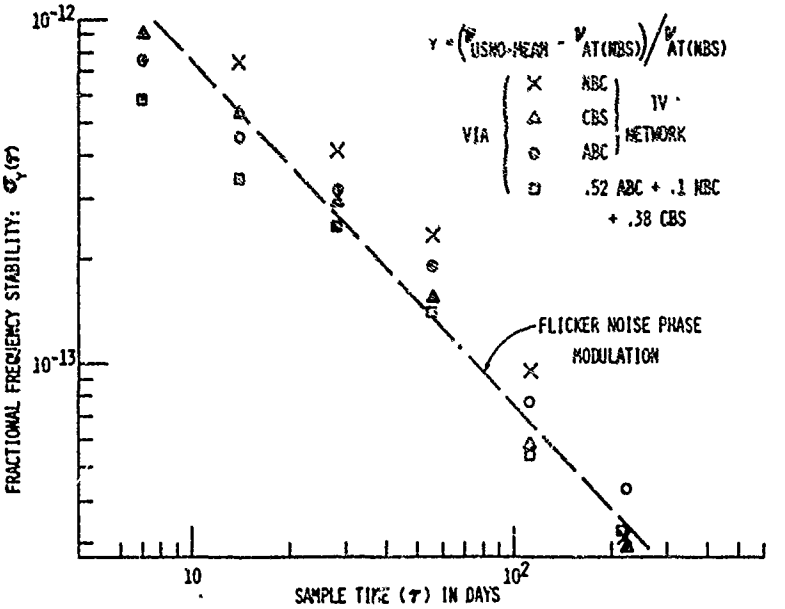


Fig. 4 - Fractional frequency stability, $\sigma_y(\tau)$, of the AT(USNO-Mean) - At (NBS) time scales compared by the 3-network TV line-10 technique.

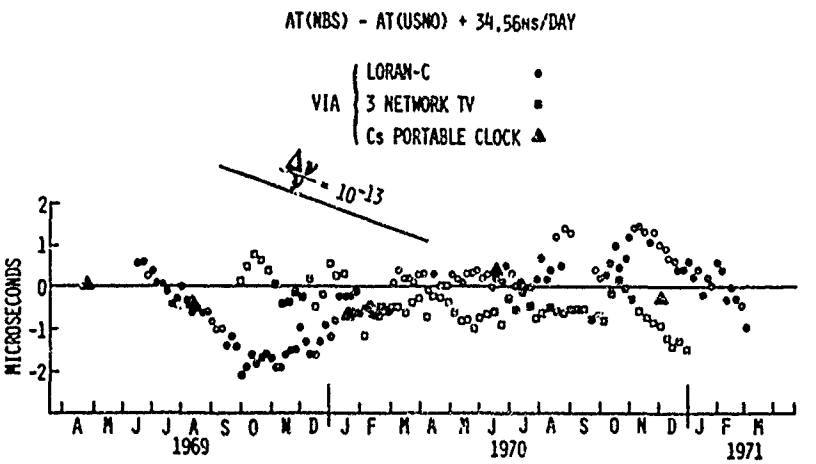
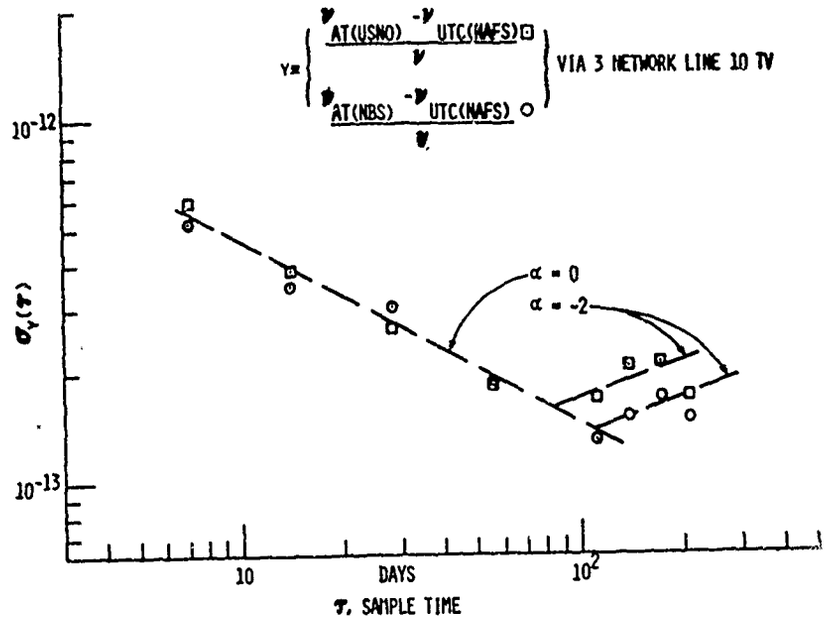


Fig. 5 - Relative time fluctuations of the AT(NBS) - AT(USNO-Mean) time scales compared by the Loran-C, 3-network TV line-10, and cesium portable clock technique.



Fig, 6 - Fractional frequency stability, $\sigma_y(\tau)$, or the Boulder-Newark and the Washington, D.C.-Newark paths via the 3-network TV line-10 technique.

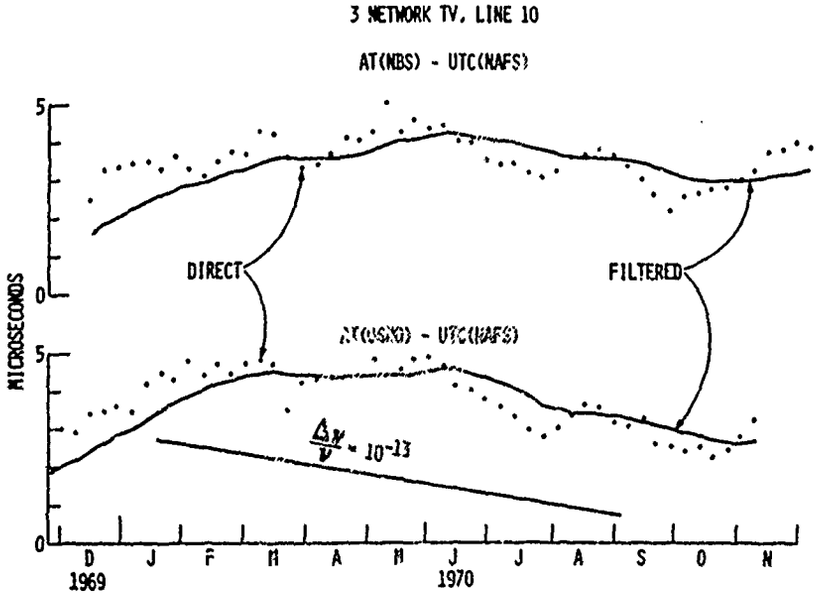
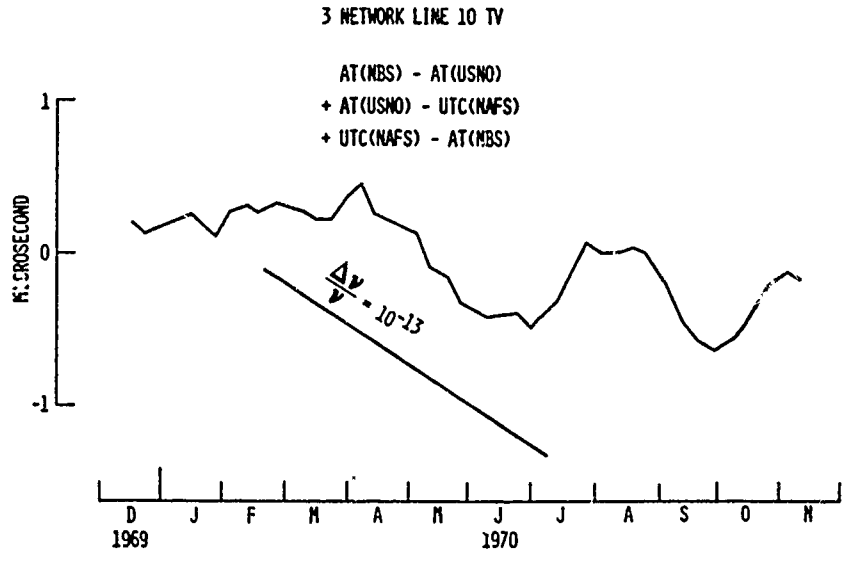


Fig. 7 - Time fluctuations (direct and filtered) of the Boulder-Newark and the Washington, D.C.-Newark paths via the 3-network TV line-10 technique



Fig, 8 - Residual time fluctuations of the 3-network TV line-10 technique between Boulder, Newark, and Washington, D.C.

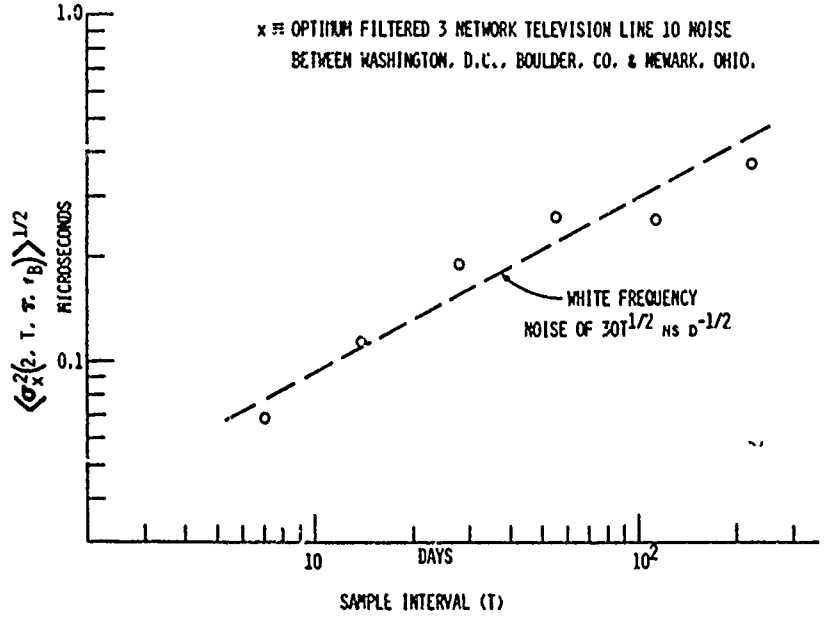


Fig. 9 - The rms time error versus sample interval of the filtered 3-network line-10 noise between Washington, D.C., Boulder, Colorado, and Newark, Ohio.

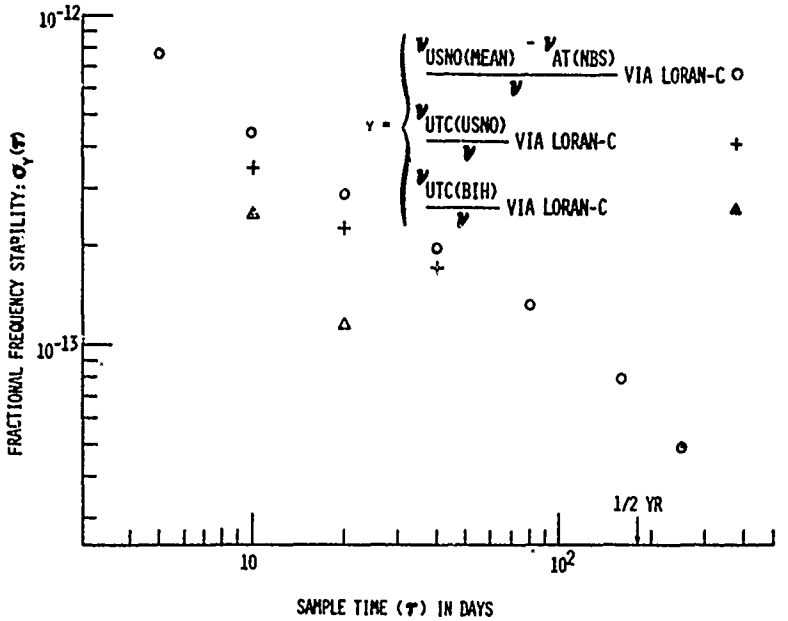


Fig. 10 - Apparent Loran-C stability: over continental U.S. path denoted by the circles; over North Atlantic path denoted by pluses; and as received at ON and BIH denoted by the triangles.

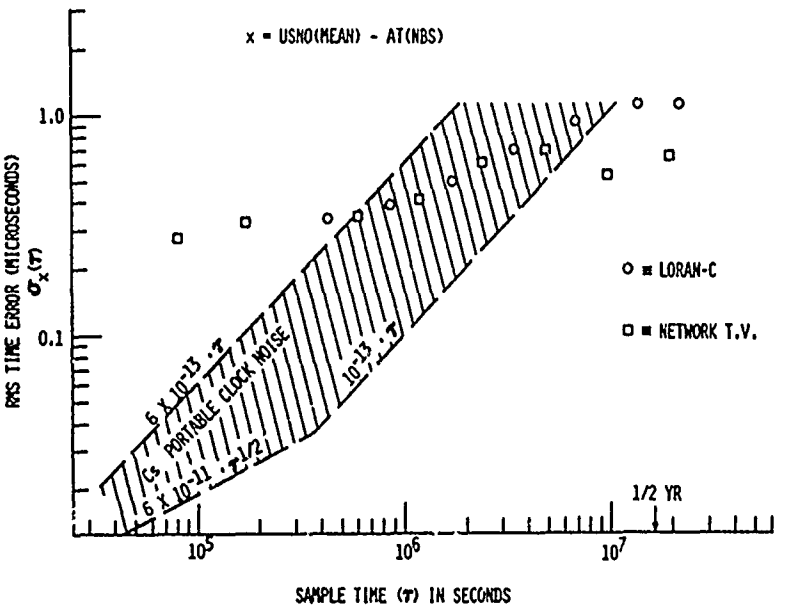


Fig. 11 - Estimation of the rms time dispersion versus sample time for Loran-C 3-network TV line-10, and cesium portable clock techniques with USNO and NBS as the time references.

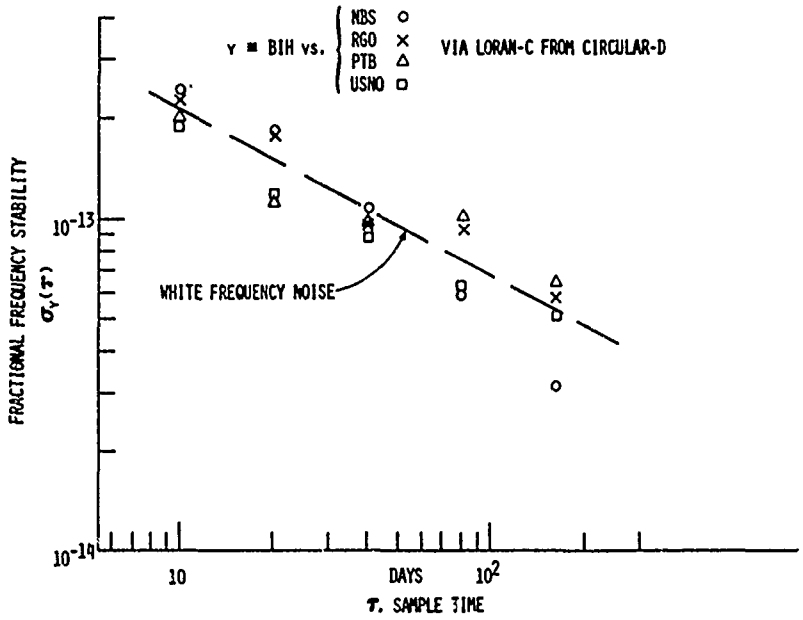


Fig. 12 - Apparent Loran-C stability as observed at the BIH from NBS, RGO, USNO, and PTB.

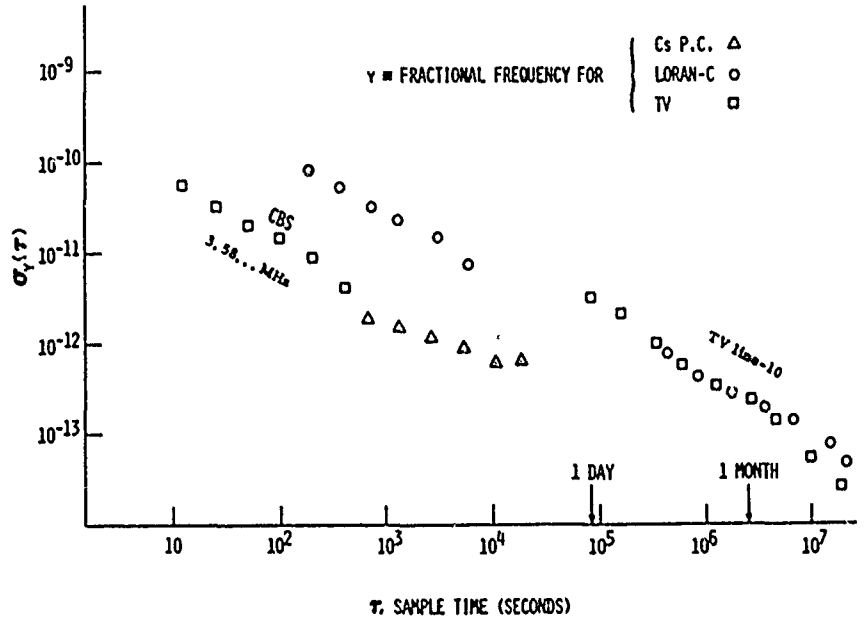


Fig. 13 - Relative fractional frequency stability versus sample time for Loran-C, 3-network TV line-10, CBS TV color subcarrier, and cesium portable clock.

METHODS OF LOCAL TIME AND FREQUENCY TRANSFER

D. H. Phillips, R. E. Phillips, J. A. Bowman, and J. J. O'Neill
Naval Research Laboratory
Washington, D. C.

Summary

Several methods of local time and frequency transfer have been experimentally utilized by the Naval Research Laboratory during the past year. Results have shown that several of the methods such as the microwave link and T.V. time transfer have great promise. Cable can provide accurate time and frequency transmission for short distances such as the Laboratory. A disciplined time and frequency standard (DTF) with built-in memory logic and a high quality crystal is being developed and could serve as the terminus of a microwave or T.V. link. The hydrogen maser was used as a transfer standard to make cross checks for various standards and between the various techniques for time transfer. Total evaluation shows the combination of all these techniques yielding a time and frequency distribution net of high accuracy to serve as a base for navigation, timekeeping, and communications.

Introduction

This paper is concerned with several of the techniques used to transfer time and frequency in a local area. First the individual systems will be described and then a description of techniques for the utilization of the systems into a complementary time transfer network will be given. Results have shown that several of the methods have great promise and can meet the needs of the user to exceedingly high accuracies.

Microwave Link

One of the most promising methods is a microwave link as shown in Figure 1 and reported on last year at the Frequency Control Symposium by D. Phillips, R. Phillips, and J. O'Neill.[1]

Additional experiments have been carried out such as higher accuracy achieved with the transmission of a 5 Mhz maser derived signal as shown in Figure 2. Full scale is 100 nanoseconds. This shows a resolution of 1 nanosecond - and it was noted that the phase noise was less than 100 picoseconds.

Portable microwave gear shown in Figure 3 was used to demonstrate four other links. The beam width of the antennas is two degrees. Another experiment accomplished and in operation is the remote controlling by the Naval Research Laboratory of power split between two microwave links utilizing the same transmitter at the U.S. Naval Observatory. This allows one to make optimum utilization of the microwave system.

The microwave links have been utilized for the comparison of remote clocks with UTC (USNO), the U.S. Naval Observatory's computed average of about 16 selected cesium beam clocks.[2] Figure 4 is supplied by the U.S. Naval Observatory and yields the fractional frequency offset and sigma values for the Hewlett Packard 5061A cesium beam 208 located at the Waldorf Microwave Space Research Facility of the Naval Research Laboratory. The top graph shows the frequency offset from 1 August 1970 to 9 March 1971 where full scale is 6 parts in 10^{12}. The lower graph shows the standard deviation in nanoseconds of a 1 day averaging time of readings taken every 3 hours and full scale is 32 nanoseconds.

The microwave link has yielded nanosecond phase resolution, tick jitter of 22 nanoseconds tick to tick or average error over 1 minute of 5 nanoseconds, and time transfer to better than 0.1 microsecond. Advantages of the microwave link are:

a. epoch time is transmitted without large unknown delay
b. wide bandwidth enables high precision transmission
c. small delay in equipment (0.4 microsecond)
d. simplicity of repair and operation
e. low transmitter power required due to point to point operation with large antenna gain

T.V. Time Transfer

Figure 5 shows another promising technique of T.V. time transfer as reported in a paper last year at the Symposium by J. L. Jesperson.[3] A cooperative experiment involving the Naval Research Laboratory, the U.S. Naval Observatory, the National Aeronautics and Space Agency, National Bureau of Standards, and WTTG, Metromedia Television in Washington, D.C. has been conducted utilizing a time code receiver designed by Dick Davis of the National Bureau of Standards and built by Astro Engineering, Boulder, Colorado. T.V. monitoring of a cesium beam at WTTG was carried out by Naval Research Laboratory from June 10, 1970 to January 11, 1971 to test the reliability and accuracy of the system. With a good antenna the system yielded nanosecond resolutions and proved reliable over this period of time.

Figure 6 shows a traveling van which was outfitted at Naval Research Laboratory to conduct moving experiments which required portable AC power. A time transfer experiment was conducted by personnel from the U.S. Naval Observatory, Naval Research Laboratory and the University of Virginia. The van, equipped with T.V., time code receiver, antenna system with amplifiers and cesium beam traveling clock, was driven to the Leander McCormick Observatory on Fan Mountain (elevation approximately 1800 ft.) on July 10, 1970. A signal was received and reliable readings of a delay of 613.385 μsec were obtained indicating a calculated distance of 114.34 statute miles. The distance calculated from coordinates was 114.8 statute miles. If we assume the distance calculated from coordinates is correct the delay was less than it should be by 0.46 statute miles or the clock in the van was ahead

by 2.47 μsec. Clock comparisons before and after indicated the clock in the van was 3.0 ± 0.1 μsec ahead of the clock at WTTG. This gave a closure within 0.6 μsec. If two T.V. stations' signals were locally synchronized and then received by equipment of this type, rho rho navigation of high accuracy would be possible for a moving vehicle in the metropolitan area. Advantages of the T.V. link are:

a. wide spread distribution of time due to wide coverage of T.V.
b. use of digital circuits make complex approach feasible
c. economical for user
d. portability

Cable

Figure 7 shows the network of cable used at Naval Research Laboratory for the transfer of maser derived frequency signals around the Laboratory. The green line is RG 223 double jacket, double shielded cable run in telephone ducts and used to transfer the maser derived 100 kHz signal around the Laboratory. Each building has an isolation amplifier. The blue line is 3/8" styroflex and is used to transfer the maser derived 5 MHz signal. The red line is 7/8" styroflex and is used to transfer a maser derived 1 MHz signal and a 1 pps signal from NRL's clock algebraically added together. For dissemination of highly accurate time the use of styroflex or other phase stable cable is recommended. Cable can provide accurate time and frequency transmission for short distances such as the Laboratory.

Correlation Using Transfer Standard

All standards involved in the various time transfer techniques are phase recorded with respect to the hydrogen maser (Naval Research Laboratory's frequency standard). This allows cross-checks to be made for the various standards and between the various techniques for time transfer. The Naval Research Laboratory's hydrogen maser is sent to the U.S. Naval Observatory via microwave link and provides a base line of low noise and high purity against which other standards can be phase compared. Figure 8 shows a phase recording of Naval Research Laboratory's two hydrogen masers beating against each other. Full scale is 700 picoseconds and each small block is 14 picoseconds. From 4 p.m. to 6 p.m. for 5 minute sampling the data yields a frequency difference of 3.2 parts in 10^{15} and a standard deviation of 3.5 parts in 10^{14}. This demonstrates the low noise characteristic of the hydrogen maser signals.

Since this high purity signal is also available throughout the Naval Research Laboratory, standards in the Laboratory can be correlated with high accuracy to the computed average of the U.S. Naval Observatory. Figure 9 illustrates local synchronization for the Laboratory.

Figure 10 illustrates data utilizing the maser as a transfer standard, the microwave link and the T.V. link. Three standards, the Cesium beam at WTTG via T.V. link (*), the U.S. Naval Observatory's Master Clock via microwave link, and the cesium beam 236 serving as Naval Research Laboratory's station clock (+) were phase recorded against NRL's hydrogen maser. The maser was common to all and could be subtracted out and thus the Cesium beam at WTTG and cesium beam 236 could be plotted against the U.S. Naval Observatory's master clock. On October 27, 1970 a C field adjustment was made to the cesium at WTTG to set it on with the master clock.

Figure 11 shows the next month's data and illustrates how useful this information was for setting and maintaining a remote clock's time.

Stabilization of Color Burst Frequency

A meeting of the T.V. Networks and FCC on August 11, 1970 at WTTG yielded a need in the TV industry for precise frequency and phase control of the 3.5795454 MHz signal utilized in color T.V. receivers to phase lock a crystal oscillator in the T.V. set. Precise measurements have been and are being made at Naval Research Laboratory to determine the changes in frequency and phase of the color subcarrier frequency of several T.V. channels. Figure 12 shows the color T.V. sets and the cable T.V. transmission system presently in use since the termination of active T.V. transmission in January.

Figure 13 shows a phase recording of WTTG Metromedia Television's Color Burst Frequency vs. Maser (NRL). Full scale is 279 nanoseconds. One can see easily the change between black and white and color. The line near the middle of the chart represents black and white operation where essentially no phase information is available. The line further on the right represents color information and illustrates stabilized color burst frequency.

Figure 14 shows WTOP's color burst frequency vs. Maser (NRL). Full scale is 279 nanoseconds. The data from 9 a.m. to 10 a.m. shows the color burst frequency going low by 3.9 parts in 10^{10} and the data from 11:30 a.m. to 11:50 a.m. shows the color burst frequency going high by 7.2 parts in 10^{11}.

Figure 15 shows WTOP's color burst frequency vs. Maser (NRL). WTOP's local color burst frequency was derived from the cesium beam standard at WTTG which had been set on frequency in October 1970 as previously shown in Fig. 10. The line from 4 p.m. to 6 p.m. shows very stable and accurate transmission of local color burst frequency. The data from 1.30 p.m. to 4:00 p.m. shows the frequency offset of the network frequency standard. Full scale is 279 nanoseconds. The stabilization of carrier frequency is a prime goal once standard frequencies are available to the T.V. station. Fringe areas receiving two stations would have minimum interference if the two stations were closely synchronized.

Disciplined Time and Frequency Standard

Figure 16 shows a model of a disciplined time and frequency standard (DTF) being developed under contract for the Naval Research Laboratory by Frequency Electronics, Inc. This standard will be readily portable, contain built-in memory logic, and contain a high quality crystal oscillator.

Figure 17 shows that the DTF will be of modular construction. Many uses presently are visualized for such a standard; for example, the terminus of a microwave or T.V. link. Another application of the DTF is when the user is separated from a precision standard by a cable. The DTF functions as a clean up oscillator when the cesium beam and cable are functioning properly. When either the cesium beam or cable fails, the DTF continues oscillating at the last available frequency - its drift at this point determined by a high quality crystal.

Conclusions

After examining these various methods of time transfer it was concluded that they are complementary

as shown in Figure 18. The results have shown that a user of precision time does not necessarily need a highly precise clock in his lab to maintain precision time. A less precise clock could be kept on time by the use of one of the described time transfer techniques. The combination of all these techniques yields a time and frequency distribution network of high accuracy to serve as a base for navigation, timekeeping, and communications.

Acknowledgements

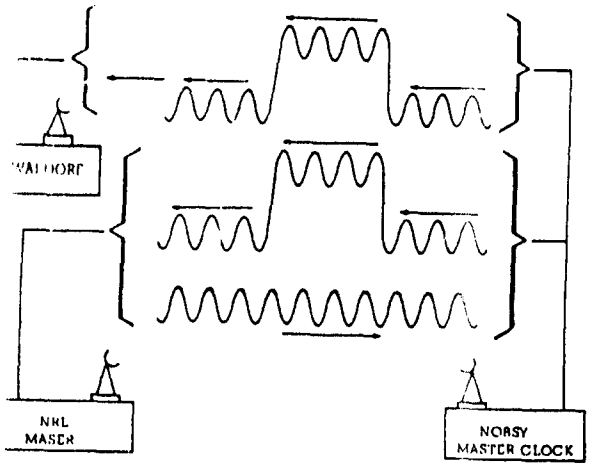
The authors would like to acknowledge with gratitude the assistance and support of the following people: Dr. G.M.R. Winkler, Mr. Harold Acrivos. Mr. D.B. Percival, and Mr. Oscar Sessions of the U.S. Naval Observatory, Mr. Dick Davis of the National Bureau of Standards, Mr. Malcomb Burleson and Mr. Albert Harmon of WTTG, Metromedia Television, Mr. Ralph Malaska of WTOP, Post Newsweek Station, Mr. R.R. Stone and Mr. T.L. Leonard of Naval Research Laboratory, Mr. Walter D. Goring and Mr. Ted Lieberman of Naval Electronics Systems Command.

References

1. D.H. Phillips, R.E. Phillips, and J.J. O'Neill "Time and Frequency Transfer Via Microwave Link", Proceedings of the 24th Annual Symposium on Frequency Control, pages 325-331, 1970. Published in IEEE Transactions on Instrumentation and Measurements, Vol. IM/-20 No. 1., February 1971.

2. G.M.R. Winkler, R.G. Hall, and D.B. Percival, "The U.S. Naval Observatory Clock Time Reference and the Performance of a Sample of Atomic Clocks", International Journal of Scientific Metrology, Vol. 6, No. 4, October 1970.

3. J.L. Jesperson, "A Survey of Time and Frequency Techniques", Proceedings of the 24th Annual Symposium on Frequency Control, pages 322-324.



Equipment: Raytheon Television Microwave Relay Model KTR-1000G

Naval Research Lab	Naval Observatory
Trans. freq. - 7137 MHz	Trans. freq. - 7415 MHz
Power - 1 watt	Power - 1 watt
Emission bandwidth - 20F9	Emission bandwidth - 20F9
Transmit hydrogen maser derived 1 MHz signal	Transmit 1 MHz and 1 pps from Master Clock

Results

Total time delay - 40.6 microseconds. (0.5 microseconds cable delay at Naval Observatory, 0.4 microseconds delay in transmitter and receiver, 0.4 microseconds cable delay at NRL. 39.3 microseconds as time traveled in air. This is equivalent to 7.3 miles.)
Precision - 2.3 nanoseconds for one day average.
Phase accuracy - 10 nanoseconds.
Tick jitter - 22 nanoseconds tick to tick (average error over 1 minute was 5 nanoseconds).
Time transfer to better than 0.1 microseconds have been achieved.

Figure 1. MICROWAVE LINK

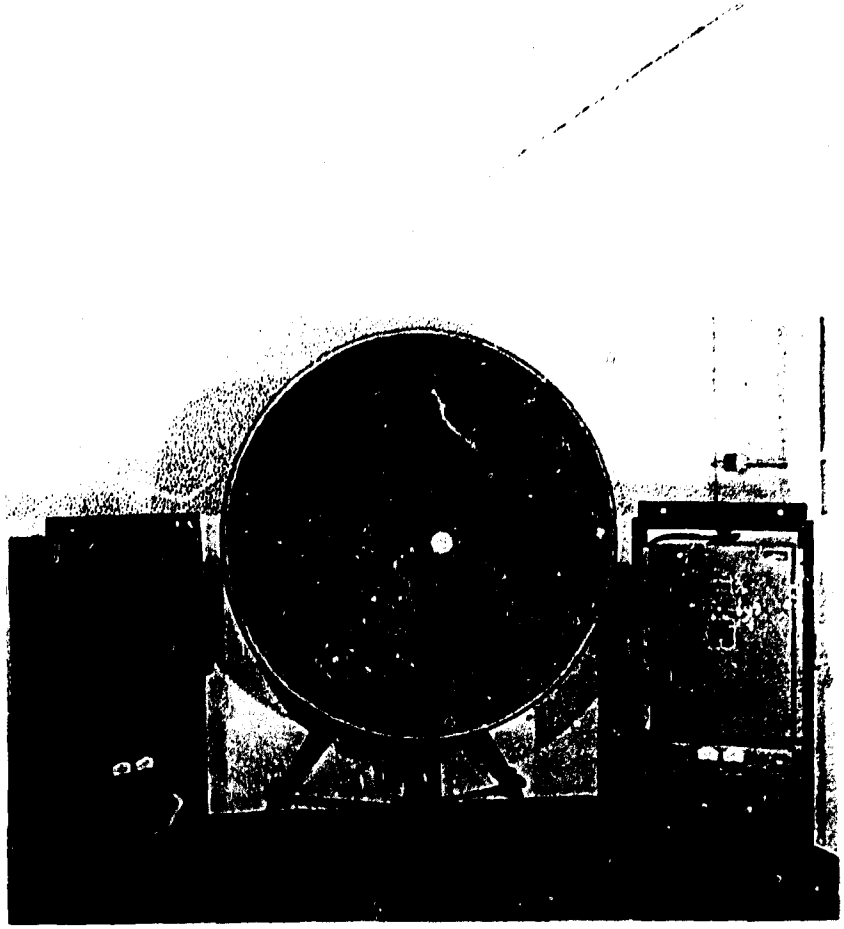
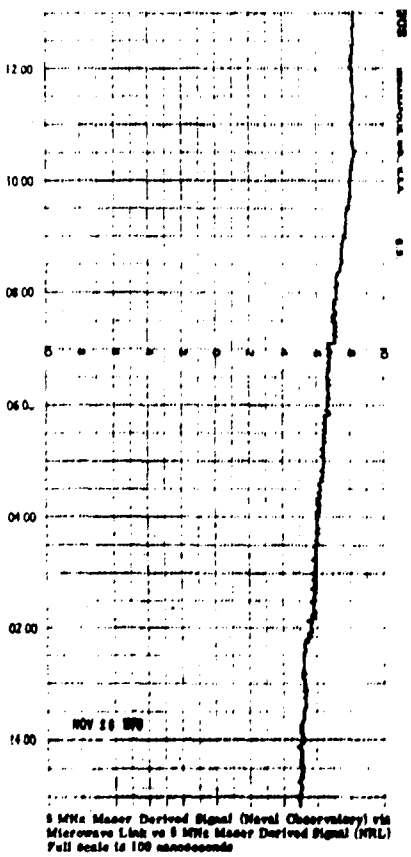


Figure 3. Portable Microwave Gear



5 MHz Maser Derived Signal (Naval Observatory) via Microwave Link vs 5 MHz Maser Derived Signal (NRL)
Full scale is 100 nanoseconds

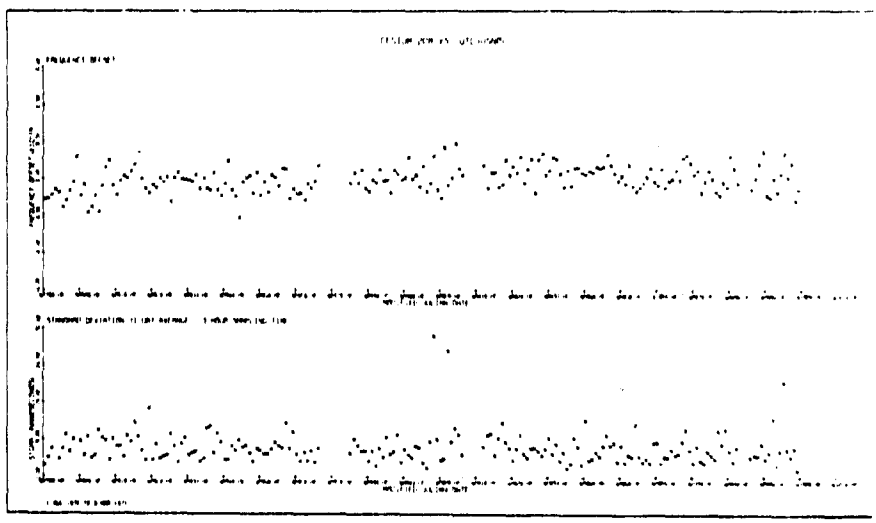
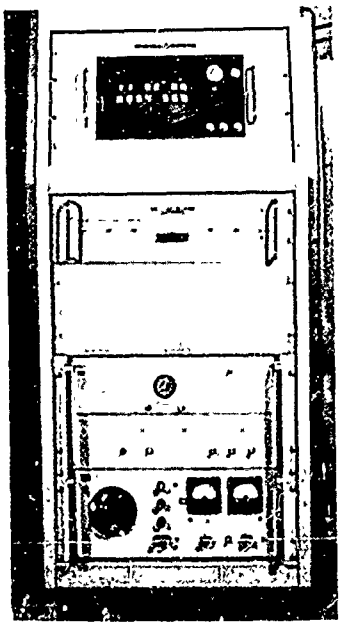
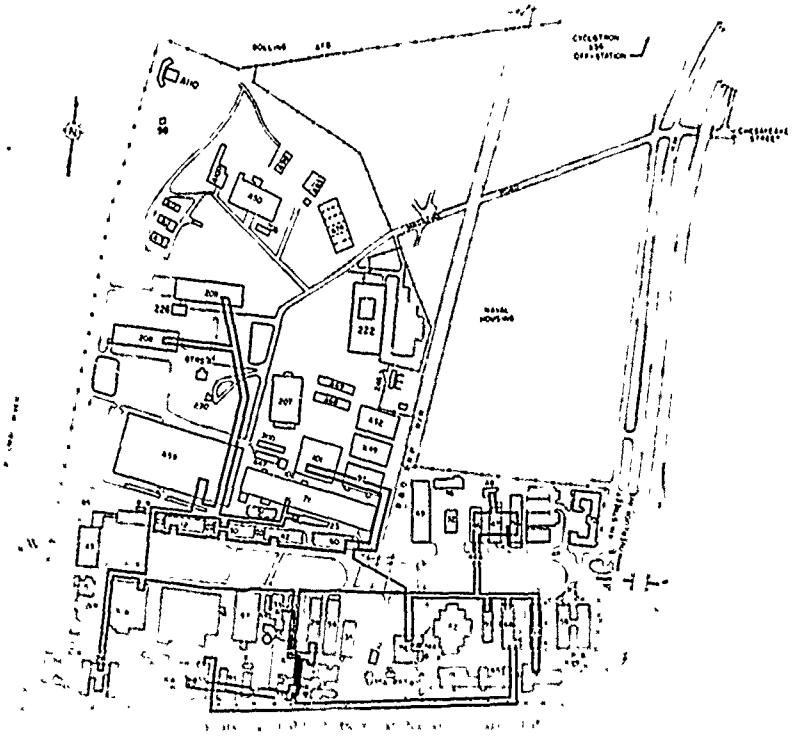
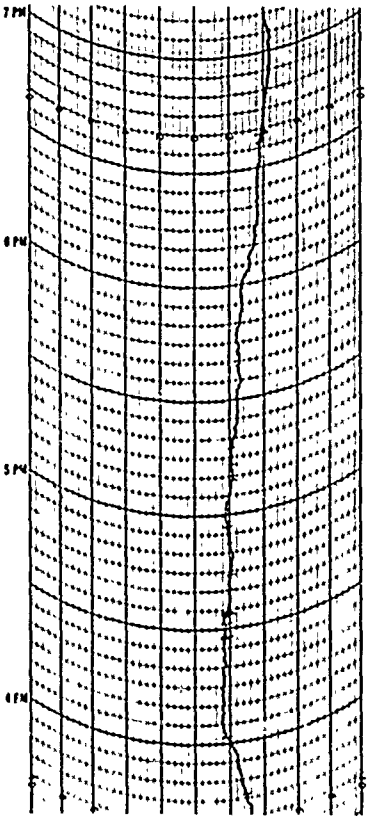


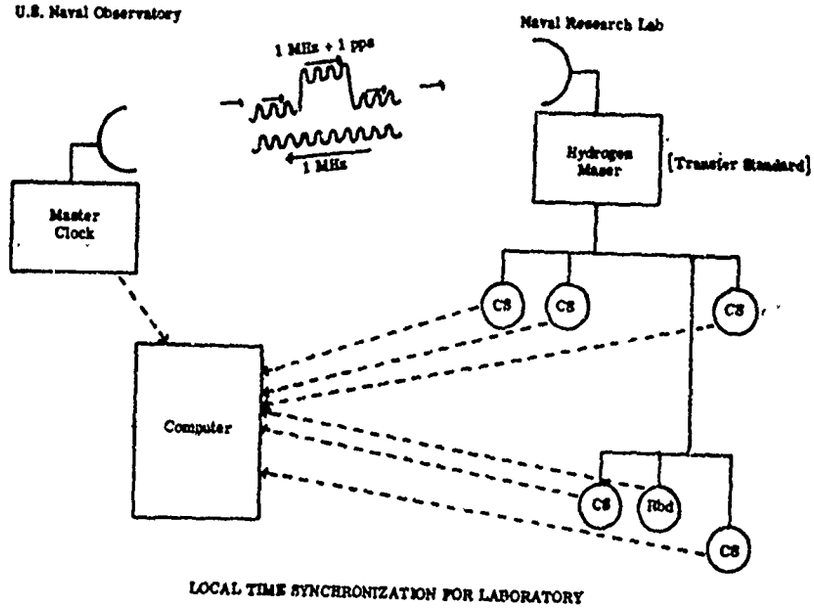
Figure 4. Remote Cesium vs. UTC (USNO)







MASER I vs MASER II
Full scale is 700 picoseconds



LOCAL TIME SYNCHRONIZATION FOR LABORATORY

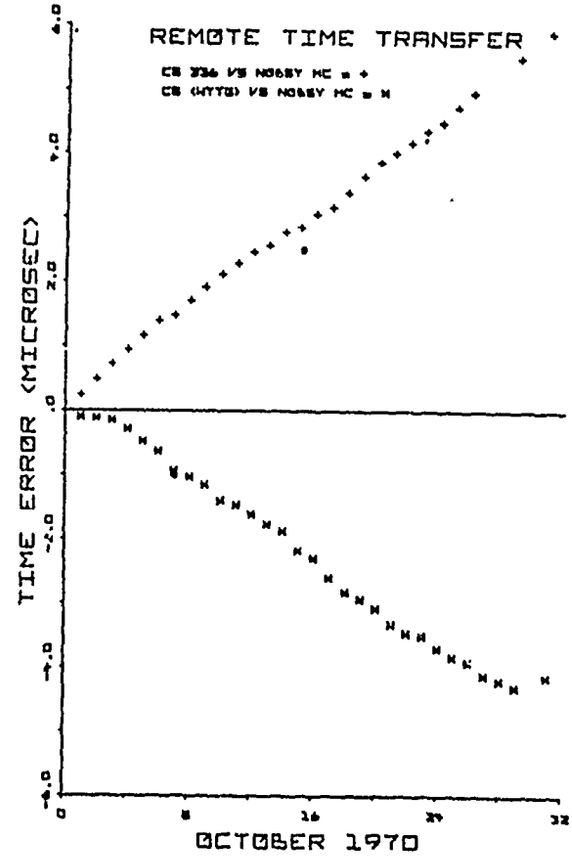
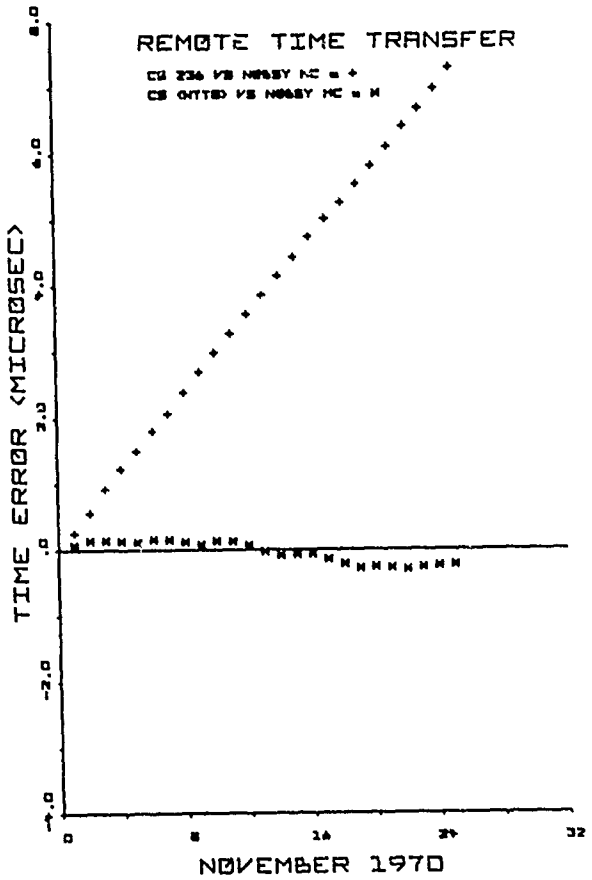
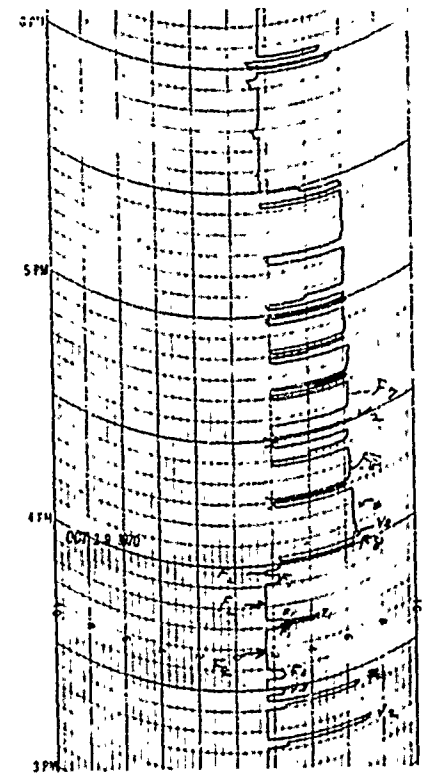


Figure 10. Remote Time Transfer (Before Frequency Set)





WTTG Color Burst Frequency vs Maser (NRL)
Full scale is 279 nanoseconds

Figure 11. Remote Time Transfer (After Frequency Set)

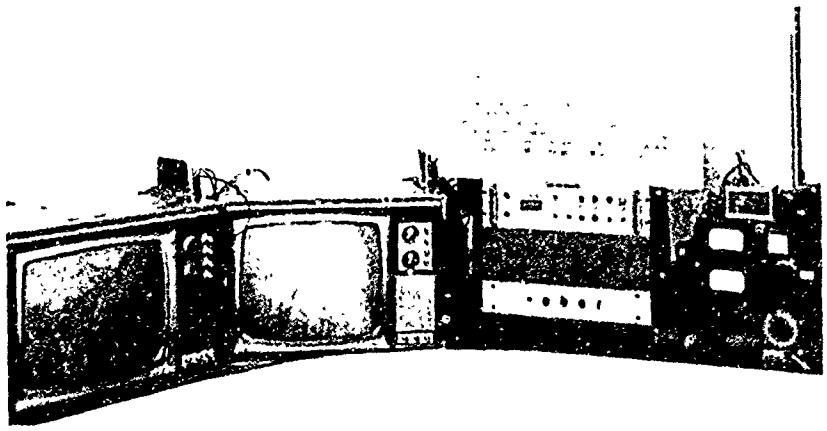
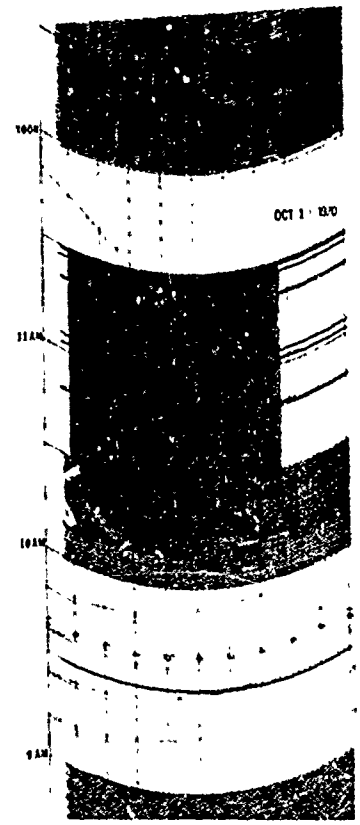
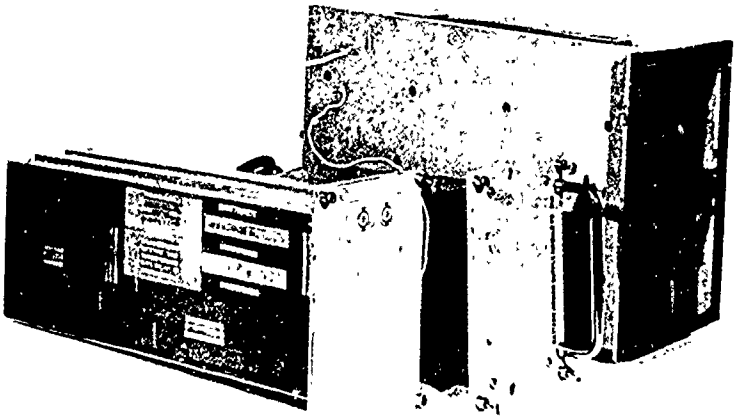
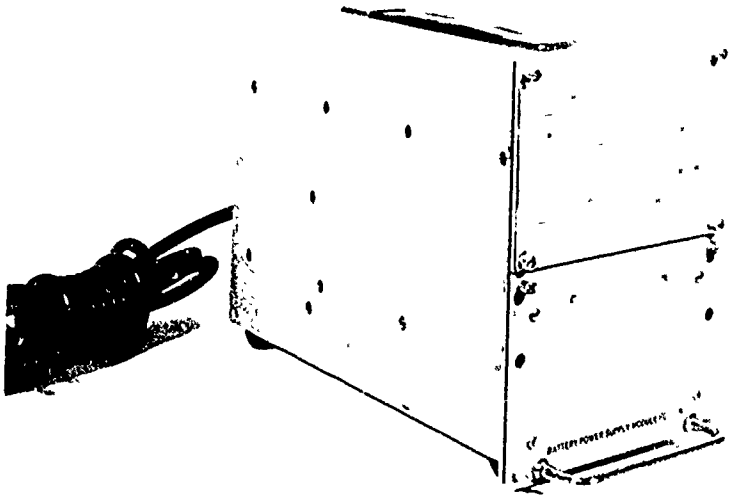


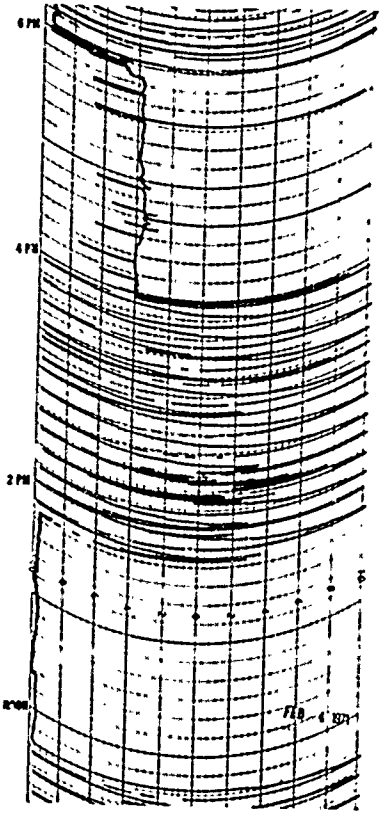
Figure 12. Color T.V. Sets and Cable T.V. Equipment



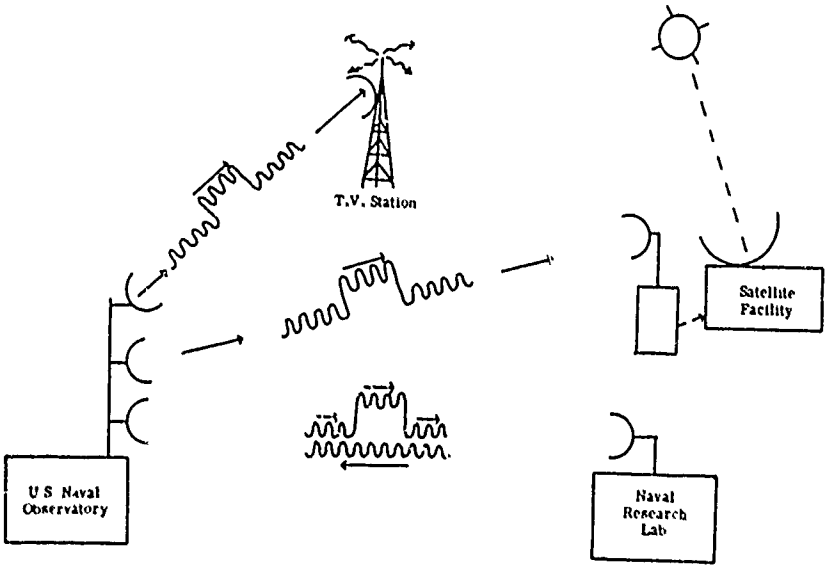
WTOP C[illegible] Burst Fre[illegible] vs Maser







WTOP Color Burst Frequency vs Maser (WTOP Local Color Burst Frequency Derived from Cesium Beam Standard at WTTG)



COMPLEMENTARY LOCAL TIME TRANSFER METHODS

INTERNATIONAL COORDINATED CLOCK TIME AND THE COMING IMPROVEMENTS IN THE SYSTEM "UTC"

G. M. R. Winkler
U. S. Naval Observatory
Washington, D. C.

Abstract

Clock time is defined by epoch and rate. Reference to an atomic resonance produces a local or "proper" atomic time scale. For all practical purposes, it is necessary to consider the following complications:

1. Clock Dispersion. One clock does not define a time scale. A procedure must be given to construct a time scale without dependence on particular clocks. On the non-local level this requires COORDINATION.

2. Relativity. Space-time geometry is not strictly Euclidic. This necessitates the concept of a "coordinate" time vs. proper time with the concept of "origin."

3. Linkage to other time scales, particularly to UT. This requires a further compromise with unpredictable adjustments.

Many users are represented by various organizations with interests and/or responsibilities in timekeeping:

a. The scientific unions under ICSU and UNESCO: IAU, IUGG and URSI.
b. The treaty organizations: CCIR (ITU), CIPM (Treaty of the Meter).
c. The "Executive Agent" Bureau: BIH.

Their functions are discussed. Coming improvements, upon which these groups have decided, will be the dropping of the "offset," one second step adjustment and a "ΔUT" code to be implemented 1 January 1972. Details are given.

Clock Time

D. H. Sadler, H. M. Nautical Almanac Office, has proposed the more general use of this term in order to provide a clear and necessary distinction between time as defined by a sequence of events in relation to a theory (ET, UTO, etc.) and clock time which is derived in a process of integration. Time is determined after the fact; clock time is directly available. Time is uniquely determined with low precision; clock time is available with very great precision but it cannot be defined without some degree of arbitrariness. Nevertheless, clock time has always been the vehicle for coordination of activities and the measurement of all other "times". We are here only interested in clock time. Time enters our picture, however, in relation to users' needs and acceptable tolerances.

Any two clocks, even the best, diverge on the average, at least with the square root of the observation interval. In addition, modern atomic clocks have operated, uninterrupted, only a few years. At the U. S. Naval Observatory, we define clock time (both UTC and A.1) as a computed average of the readings of a group of independently operating cesium beam clocks. Procedures have been established to insure that clocks can be added or replaced, ideally, without introducing steps in time or rate. We feel that this is a great improvement over the procedures used from the beginning (1956) and it is a method which can be extended over any set of timekeeping centers as was proposed by B. Guinot. It is the only method which allows the production of a most uniform clock time with maximum reliability locally and, with some degree of degradation due to the effects of measurement delay times and noise upon the "control loop", it also allows the production of an improved internationally coordinated time.

At this point it becomes clear that there is a significant practical difference between the operation of a modern atomic clock and a frequency standard. Since the atom cannot be observed without any disturbances and errors, it turns out that one has a choice:

a. To keep the disturbances small but constant, or
b. To determine the disturbances frequently in a process of calibration.

a. represents the operation of a clock; b. represents a frequency standard. In our experience the interests of timekeeping (uniformity, reliability, redundancy) require procedure a. Procedure b. is necessary to determine the "true" frequency of the atoms of interest (cesium, thallium, hydrogen, etc.). In the past we have adjusted the basic frequency for clock time operations very infrequently. For instance, the average frequency used since 1968 has not been changed noticeably at the U. S. Naval Observatory. All improvements introduced affected merely the random walk of the "Master Clock" used (reference (1)). It is clear from such experience that one can today derive time scales approaching a uniformity of about 100 nanoseconds in 100 days (10^{-14}) using ordinary but well selected commercial cesium beam clocks, properly operated, as a "reference assembly" without knowing where the actual frequency of cesium is within one part in 10^{12}.

Coordinated Coordinate Time

Because of random walk, all clocks diverge from a mean time scale. In addition, the intercomparison

between distant clocks cannot be done continuously, and the delays and the measurement (phase) noise involved degrade the uniformity of an average of widely spatially distributed clocks. A good compromise is a hierarchy of coordinated clocks, all of which must make small adjustments from time to time in order to remain within specified tolerance limits of the system.

The group tolerances can be smaller the further down one goes into the "local" level. At the present time we experience the need for about four (4) such levels: the international level (BIH, level (1)); the National Time Services, level (2); which, in turn, is composed of a series of major time using and distributing centers, level (3); and finally the user clock level which is often structured even further. The National Time Services should be responsible for coordination of activities. Currently our coordinated time (UTC) specifies 1 millisecond clock time tolerance and few time services have had to make small steps in the past few years. With presently available clocks, the tolerance could easily be reduced to 100 microseconds for the level (2) and to 10 microseconds within the level (3) groups. A gradual tightening of tolerances is recommended because of the need for more experience at the working level. After several years, these tolerances could be further reduced.

In the United States and in France, examples for high precision coordination exist. In the interest of improved coordination between the National Bureau of Standards (NBS) and the U. S. Naval Observatory (USNO), all measurements at the USNO are referred to a coordinated clock known as the USNO Master Clock and designated as MC(USNO). This clock differs by very small amounts from the internal (Local) USNO clock time reference which is designated Mean (USNO). The frequency difference, during 1969 for example, was as follows:

$$\frac{F(MC) - F(Mean)}{F(Mean)} = 0 \text{ before 1 May 1969}$$

$$\frac{F(MC) - F(Mean)}{F(Mean)} = -1 \times 10^{-13} \text{ after 1 May 1969.}$$

For most timing applications these changes are insignificant since they are less than the random fluctuations of individual atomic clocks. By agreement, opposite frequency adjustments have been introduced by NBS. The coordinated clocks of these two agencies have been within 5 microseconds since July 1968. The internal (local) time scales serve as a basis for the Bureau International de l'Heure (BIH) time scale. The coordinated clocks are used for all external measurements. The adjustments are published in Time Service Bulletins, Series 11.

Another aspect of systems of widely distributed clocks, as yet relatively insignificant, must be carefully considered in view of new technological developments. This aspect is twofold and should be described without reference to a particular theory (general relativity) but need only refer to the two effects which are on safe experimental grounds and which can be considered separately for our purpose:

1. The gravitational "red shift".

2. The transversal Doppler effect (Lorentz Transformation).

The frequency of a moving clock is increased by a relative amount of approximately

$$\frac{\Delta F}{F} = \frac{1}{c^2}\left(\Delta\Phi - \frac{1}{2}v^2\right),$$

where $\Delta\Phi$ is the difference in gravitational potential and c is the velocity of light. This amounts to about 1.1 part in 10^{13} for an altitude difference of 1 km at rest. If the clock moves in a jet plane at 10 km altitude and at a speed (v) of 1000 km per hour, the effect is about $+ 6 \times 10^{-13}$ due to both terms. Our present flying clock operations are just at the point where these small corrections will have to be applied. The effect can be different, however, depending on the direction of the flight (reference (2)).

In other words, for terrestrial purposes we have to specify a reference level of gravitational potential, and it appears natural to use sea level as the reference or "origin" for future "coordinate clock time" systems (time is used here as a "coordinate"). The BIH in Paris has actually been adopted as reference. One can ignore, at present, the annual variation in the rate of atomic clocks due to the ellipticity of the orbit of the earth-moon system around the sun since it is of no practical consequence for operations on the surface of the earth. Astronomers, however, are concerned with this effect (annual variation is about 2×10^{-10}) for example, in pulsar observations (J. G. Duthie and Paul Murdin, University of Rochester). See also reference (3).

Difference Between Time and Time Interval

In order to have any widespread usefulness, a time scale must suffer some adjustments or compromises in the interest of its applicability in nature. A time scale cannot be compared with other physical scales as length, temperature, etc. This is the root of many misunderstandings. Rather, a time scale can be compared to a coordinate system similar to the geodetic grid used on the surface of the earth.

In analogy to length measurements, time interval can correspond to the distance between two locations, a and b, on the surface of the earth. This can be measured, either directly or by reference to geodetic or astronomical coordinates of a and b. In the second case a computation has to be performed taking into account the non-uniformity of the coordinate system on the surface of the earth due to the oblateness of the earth's spheroid. Similarly, time scales are in existence corresponding to such coordinate systems which may or may not be uniform but do allow the measurement of time intervals by applying well known corrections.

If one measures distance over short intervals, then one should do it directly in relation to the standard meter and not by giving the coordinates of the two points. On the other hand, we accept it as entirely practical to compute very great distances from the coordinates even if the grid is not uniform

but distorted (spheroid) as long as we know that distortion. The situation here too, is a compromise in deference to the same cause -- our "imperfect" earth.

Very similarly, if we insist on using the global international timing system for laboratory "absolute" frequency calibrations instead of using the standard of frequency, cesium atoms in the local frame of reference, then we have to pay the price of applying small but precisely known corrections.

In summary the following points are important:

A hierarchical system of coordinated time, existing now in a rudimentary stage, can be developed. Corrections for "relativity" effects are small and can be applied easily when required. Small adjustments must be made regularly, necessitating the continued existence of two clock time scales at major centers:

(1) The local, proper, "atomic" time, never to be adjusted, for the purpose of the definition of clock time (A.1, Mean at the U. S. Naval Observatory).

(2) A "coordinated" time to be distributed for the realization of a generally available time scale (UTC).

Organization

The international scientific unions, URSI, IUGG and particularly the International Astronomical Union (IAU) have, in the past, been most concerned with the scientific and practical aspects of the determination and use of the various time scales including clock time. The BIH plays the role of the executive agent for the coordination and prompt publication of time-related matters. It is sponsored by the IAU and receives a little support from UNESCO through "FAGS", the Federation of Astronomical and Geophysical Services. More recently, the "Treaty of the Meter" with its various bodies CIPM, BIPM, etc., has formally defined the (atomic) second as a base unit of the "Système International" (SI) and has also formally defined an International Atomic Time Scale (IAT) for common reference purposes (in agreement with the IAU definition) to be kept by the BIH.

Based on this current interest, the BIH also acts as an agent for the CIPM in regard to the IAT.

The International Telecommunications Union (ITU), with its advisory body the International Consultative Committee for Radio (CCIR), agrees on standards for radio time signals and standard frequency transmissions including assignments of protected frequency bands, etc.

It is clear that changes in policy or procedures require a great degree of coordination and a rather deliberate, conservative approach in order to avoid vacillation and confusion. Steady progress, however, has been achieved in adapting our procedures and standards to the ever developing needs and capabilities of modern technology.

Adjustments to Universal Time

Requirements for Time

Time, in the sense "time of day", is needed as one of the fundamental parameters in astronomical and navigational tables. This, together with the requirements of civil use (scheduling, legal time, etc.), necessitates a close link of time signals with diurnal phenomena, i.e., mean solar time. On the other hand, numerous new global electronic systems demand synchronized operation with a precision of about one microsecond. Some of these systems cannot tolerate the small time steps or frequency changes required in the present UTC system where the frequency of occurrence of both is in the dangerous zone of "too often, but not frequently enough". Aircraft, for example, may soon be using microsecond timing systems which cannot tolerate these changes for either safety or logistic reasons. In view of these new developments certain improvements have been decided upon by the bodies previously mentioned. Essentially, the changes to take place 1 January 1972, 0^h UT are as follows:

1. We will discontinue the "offset".
2. We will approximate UT with a larger tolerance by stepping exactly 1 second when necessary (about once per year, preferably on the last second of 31 December or 30 June).

This method of using "leap" seconds is exactly analogous to the calendar adjustments with its leap day. Since UT1 is required by most users in astronomy, navigation and geodesy, additional information will be available on the time signals which will give the difference

$$\Delta UT = UT1 - UTC \text{ (Time Signal)}.$$

The CCIR has recommended a special code (see appendix C). The use of voice or Morse Code to provide the necessary correction for the use of time signals in observations of "Hour Angle" is also possible.

The following details will apply:

a. On 1 January 1972 0^h UT we will

(1) increase the frequency of UTC by 300×10^{-10}.

(2) delay (retard) UTC(USNO) by approximately 107,607 microseconds. (The exact amount will be announced in the fall of 1971.)

b. After these adjustments we will have:

(1) IAT - UTC(USNO) = 10.0 sec (until the next step).

(2) $\Delta UT = + 0.1$ sec (decreasing at a rate of about 2.5 ms per day).

c. The first 1 sec step can be anticipated for the end of June 1972.

We realize that the frequency change will create special problems for those users of atomic frequency standards (cesium beam, rubidium gas cell) who do not

have provisions available for such changes.

However, the new system has been decided upon for the purpose of eliminating future frequency changes. We all members of the time community to take "the steps" to make the transition easy.

ANNEX A

XIV GENERAL ASSEMBLY IAU -- BRIGHTON 1970

Commission 31 - Resolution #1 - 24 August 1970

Commission 31 makes the following recommendations:

1. That the frequency offset of UTC be made zero, effective 0^h January 1972.

2. That step adjustments shall be exactly 1^s. When a step adjustment is made it shall be at the last second of a UTC month with preference for 31 December or 30 June. These step adjustments will be decided upon and announced as early as possible by the BIH.

3. The maximum difference $|UTC - UT1|$ will be less than $0^s.7$ unless there are exceptional variations in the rotation of the Earth.

4. Special adjustment. The BIH will also announce a unique fraction of a second adjustment to be made at 0^h 1 January 1972, so that UTC and the International Atomic Time Scale (IAT, in French TAI) will differ by an integral number of seconds.

5. The emission times of time signals from coordinated stations shall be kept as close to UTC (BIH) as feasible with a maximum tolerance of 1 ms.

6. Nomenclature.

6.1 Clocks in common use will indicate the minutes, seconds and fractions of UTC (French: TUC).

6.2 The terms "G.M.T." and "Z" are accepted as the general equivalents of UTC in navigation and communications.

7. The term ΔUT is defined by: $\Delta UT = UT1 - UTC$. Extrapolated and final values of ΔUT will be issued by astronomical observatories and the BIH, and will be given the widest possible distribution.

8. All standard time signal emissions must include information which will enable a user to obtain UT1 with a precision of at least $0^s.1$.

9. Designation of the epoch of steps in UTC

9.1 If UTC is to be advanced, then second 00 will follow $23^h\ 59^m\ 58^s$ of the previous day.

9.2 If UTC is to be retarded, then the second of the previous day $23^h\ 59^m\ 60^s$ will be followed by the next second $0^h\ 00^m\ 00^s$ of the first day of the month.

9.3 The stepped second will be commonly referred to as a "leap" second (in French: intercalaire).

9.4 The time of an event given in the old scale, before the leap second, will be given as a date in the previous month, exceeding 24^h if necessary. The time of an event given in the scale after the step will be given as a date in the new month, with a negative time, if necessary.

NOTE: Commission 31, taking into account the conflicting requirements of the various users of UTC, including the large number of those requiring immediate knowledge of hour angle, consider that the above represents the optimum solution.

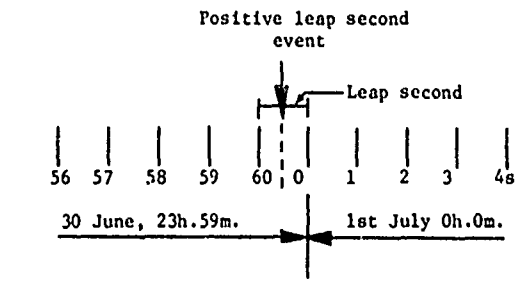
ANNEX B

(From CCIR Report 7/13E, 22 February 1971)

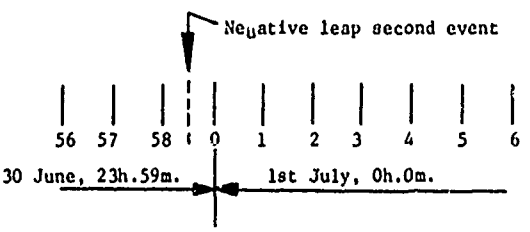
Dating of Events in the Vicinity of a Leap Second

A positive or negative leap second, when required, should be the last second of a UTC month, preferably 31 December and/or 30 June. A positive leap second begins at $23^h\ 59^m\ 60^s$ and will be followed one second later by $0^h\ 0^m\ 0^s$ of the first day of the month. In the case of a negative leap second, $23^h\ 59^m\ 58^s$ will be followed one second later by $0^h\ 0^m\ 0^s$ of the first day of the month.

Taking into account the contents of this paragraph, dating of events in the vicinity of a leap second shall be made as indicated in the figures below:



Designation of the date of the event
30 June, $23^h\ 59^m\ 60.6^s$ UTC



Designation of the date of the event
30 June, $23^h\ 59^m\ 58.9^s$ UTC

ANNEX C

(From CCIR Report 7/13E, 22 February 1971)

Code for the Transmission of DUT1

A positive value of DUT1 will be indicated by emphasizing a number (n) of consecutive seconds markers following the minute marker from seconds marker one to seconds marker n; n being an integer from 1 to 7 inclusive.

DUT1 = (n x 0.1)s

A negative value of DUT1 will be indicated by emphasizing a number (m) of consecutive seconds markers following the minute marker from seconds marker nine to seconds marker (8+m); m being an integer from 1 to 7 inclusive.

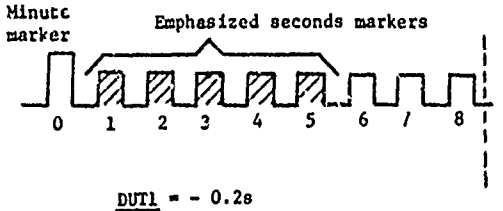
DUT1 = - (m x 0.1)s

A zero value of DUT1 will be indicated by the absence of emphasized seconds markers.

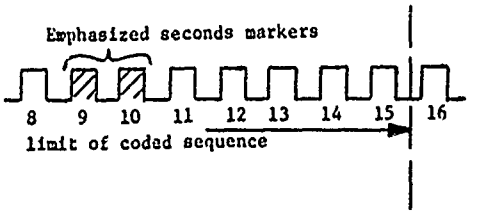
The appropriate seconds markers may be emphasized, for example, by lengthening, doubling, splitting, or tone modulation of the normal seconds markers.

EXAMPLES

DUT1 = + 0.5s



DUT1 = - 0.2s



Reference

1. G. M. R. Winkler, R. G. Hall and D. B. Percival, "The U. S. Naval Observatory Clock Time Reference and the Performance of a Sample of Atomic Clocks", Metrologia, Vol. 6, No. 4, October 1970, pp. 126-143.

2. J. C. Hafele, "Relativistic Behavior of Moving Terrestrial Clocks", Nature, Vol. 227, No. 5255, 18 July 1970, pp. 270-271.

3. G. M. Clemence and B. Szebehely, "Annual Variation of an Atomic Clock", Astronomical Journal, Vol. 72, No. 10, December 1967, pp. 1324-1326.

FREQUENCY MODULATION ANALYSIS WITH THE HADAMARD VARIANCE

Richard A. Baugh
Hewlett Packard Company
Santa Clara, California

Introduction

It has been shown that the spectral density of a signal source can be derived from time domain measurements made with a frequency counter (1) if the spectral density is of the form

$$S_y(f) = f^n \qquad n = -3, -2, -1, 0, \text{ etc. and } f < f_c$$

$$S_y(f) = 0 \qquad \text{otherwise.}$$

This involves plotting the sample variance versus averaging time and measuring the assymptotic slope. The development of frequency counters which can be programmed to make measurements and do arithmetic calculations simultaneously makes it possible to do high resolution spectral analysis on signal sources. The counter is programmed to do an operation resembling the Hadamard transform (2), similar to the finite Fourier transform but with the sine and cosine functions replaced with square waves.

Properties of the Allan Variance

Because the spectral density of frequency fluctuations almost always has a singularity at f=0 special statistical quantities such as the Allan variance must be used (1). $\sigma_y^2(2,\tau)$, the Allan variance, is defined as the mean square value of $\sqrt{\tfrac{1}{2}}(\bar{y}_k - \bar{y}_{k+1})$:

$$\sigma_y^2(2,\tau) = \left\langle \frac{(\bar{y}_{k+1} - \bar{y}_k)^2}{2} \right\rangle$$

where $\bar{y}_k$ is the fractional frequency averaged from t_k to $t_k + \tau$:

$$\bar{y}_k = 1/\tau \int_{t_k}^{t_k+\tau} y(t)\, dt$$

and $t_{k+1} = t_k + \tau$.

In order to understand the relationship between the Allan variance of a signal source and the spectral density of its frequency fluctuations a non-mathematical example can be given: If as shown in Fig. 1(a) the averaging time, τ, is equal to a half-period of the sinusoidal frequency modulation and is in phase with it then the Allan variance will be large.

If, however, the measurement time is not approximately equal to a half period or if it is in quadrature with the modulation (Fig. 1(b)), then the Allan variance will be much smaller. Thus, one would expect the Allan variance to depend primarily on modulation components near $f = 1/2\tau$.

Placing the definition of $\bar{y}_k$ inside the statistical averaging bracket the Allan variance becomes the mean square value of a convolution integral.

$$\sigma^2(2,\tau) \left\langle \tfrac{1}{2}\left(1/\tau \int_{t}^{t+\tau} y(u)\, du - 1/\tau \int_{t-\tau}^{t} y(u)\, du \right)^2 \right\rangle$$

or

$$\sigma^2(2,\tau) = \left\langle \left(\int_{-\infty}^{+\infty} y(u)\, h(u-t)\, du \right)^2 \right\rangle$$

with

$$h(t) = -1/\sqrt{2\tau} \qquad -\tau < t < 0$$

$$1/\sqrt{2\tau} \qquad 0 < t < \tau$$

$$0 \qquad \text{otherwise.}$$

The instantaneous frequency is convolved with a function h(t) which resembles the measurement sequence. Replacing the convolution with a filter whose impulse response is h(t) it is now possible to separate the various operations involved in obtaining the Allan variance.

The output of a linear system is the input convolved with h(t) or, since the Fourier transform of a convolution is the product of the two Fourier transforms,

$$S_{y_1}(f) \quad S_y(f) \cdot |H(f)|^2$$

where H(f) is the transfer function or Fourier transform of the impulse response, h(t), and $S_y(f)$ and $S_{y_1}(f)$ are the spectral densities of the input and output. Consequently, the relationship between $\sigma^2(2,\tau)$, the Allan variance, and the spectral density is

$$\sigma^2(2,\tau) = \int_0^{+\infty} S_y(f)\, |H(f)|^2\, df$$

The transfer function for the Allan variance, shown in Fig. 3, has a broad peak near $f = 1/(2\tau)$. Because this peak is so broad the Allan variance is not particularly useful for spectral analysis.

Properties of the Hadamard Variance

Another statistical quantity, the Hadamard variance, can be defined which has a much narrower transfer function, making if useful for spectral analysis of frequency modulated signals. In this method the frequency is averaged over many successive time intervals and alternately added ans subtracted. The Hadamard variance is the mean squared value of

$$|\bar{y}_{k+1} - \bar{y}_{k+2} + \bar{y}_{k+3} \quad - \bar{y}_{k+2N}|.$$

Specifically,

$$\sigma_H^2(N,\tau) = \left\langle |\bar{y}_{k+1} - \bar{y}_{k+2} + \bar{y}_{k+3} \cdots - \bar{y}_{k+2N}|^2 \right\rangle.$$

The Allan variance is one half times the Hadamard variance with N = 1.

The impulse response of the Hadamard variance is a segment of a square wave with N complete cycles. Its transfer function, the Fourier transform of the impulse response, is shown in Fig. 5. The bandwidth is 1/N times the repetition frequency since the impulse response contains N complete cycles. The Hadamard variance is useful for spectral analysis because it only responds to a very narrow range of frequencies.

One disadvantage of the Hadamard variance over conventional spectral analysis is evident from Fig. 5. Since the impulse response is a segment of a square wave (which contains odd harmonics) spurious responses are present. A frequency counter programmed to respond to frequency modulation at a frequency $1/2\tau$ would also respond to any modulation which was present at $3/2\tau$, $5/2\tau$, etc.. This can be partly eliminated by putting a delay or dead time between frequency averages. The optimum dead time is 50% of the measurement time. One can say that a square wave with 50% dead time is a better approximation to a sine wave than a square wave with no dead time. The 50% dead time eliminates the third, ninth, fifteenth, etc., harmonic responses as illustrated in Fig. 6.

The measurement sequences illustrated in Figs. 5 and 6 produce transfer functions with large sidelobes. These sidelobes can be modified or eliminated by multiplying the impulse response by an appropriate envelope function. The so-called Hanning and Hamming windows are envelope functions used for the same purpose in digital Fourier analysis (3). An example of an envelope function for the Hadamard variance is given in Fig. 7. The envelope, which consists of the binomial coefficients, completely eliminates the sidelobes.

Applications

The Hadamard variance is a useful measurement method because it extends frequency domain measurements down to sub-audio frequencies and because it measures frequency modulation directly. It can be implemented with a programmable high-frequency counter such as the Hewlett-Packard 5360A, a programmer, and a square wave generator which is used to adjust the measurement sequence. The equipment, illustrated in Fig. 8, can measure low frequency FM sidebands and spectral purity, non-linearity and intermodulation in FM modulators, and with the aid of a voltage-to-frequency converter it can be used for sub-audio spectral analysis.

Appendix A

Obtaining the FM Deviation and Spectral Density from the Hadamard Variance.

In order to translate Hadamard variance measurements into more useful quantities it is necessary to know the exact mathematical relationship between the Hadamard variance and FM spectral density. It can be shown that

$$\sigma_H^2(\tau,T_D,N) = \int_0^\infty |H(f)|^2 S(f)\,df \qquad (1)$$

where S(f) is the FM spectral density,

$$H(f) = 2N\sum \operatorname{sinc}\left(\frac{n\pi}{2(\tau+T_D)}\right)\operatorname{sinc}\left[2N(\tau+T_D)\left(f-\frac{n}{2(\tau+T_D)}\right)\right] \qquad (2)$$

$$n = 1,3,5\cdots$$

the sinc function is defined as

$$\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.$$

τ is the measurement time,

T_D is the dead time between measurements, and

N is the number of complete cycles in the measurement.

The amplitude of an FM component at

$$f = \frac{1}{2(\tau+T_D)}$$

is $\Delta f_{rms} = \dfrac{\sigma_H}{A_1}$

where $A_n = 2N \operatorname{sinc}\left(\dfrac{n\pi\tau}{2(\tau+T_D)}\right)$. (3)

assuming that the phase of the FM component is uniformly distributed between 0 and 2π radians. Similarly, the amplitude of an FM component at an arbitrary frequency is

$$\Delta f_{rms} = \frac{\sigma_H}{|H(f)|} \qquad (4)$$

Since the equivalent noise bandwidth, B_N, of each peak in the frequency response is

$$B_N = 1/N = \frac{1}{2(\tau+T_D)}$$

the integral of Eq. (1) can be replaced by a sum,

$$\sigma_H^2(\tau,T_D,N) = (1/N)\left(\frac{1}{2(\tau+T_D)}\right)\sum A_n^2 S\left(\frac{n}{2(\tau+T_D)}\right) \qquad (5)$$

$$n = 1,3,5\cdots$$

For bandlimited noise which is zero for $f > \frac{1}{(\tau + T_D)}$ the relationship between the Hadamard variance and the FM spectral density evaluated at

$$f = \frac{1}{2(\tau + T_D)} \quad \text{is simply}$$

$$S\left(\frac{1}{2(\tau + T_D)}\right) = (2N)(\tau + T_D)\left(\frac{1}{A_1}\right)^2 \sigma_H^2 \tag{6}$$

If the noise is bandlimited at a higher frequency so that there is also a noise contribution from the third or higher harmonic responses of the Hadamard variance then

$$\sigma_H^2(\tau, T_D, N) = \frac{1}{2N(\tau + T_D)}\left[A_1^2 S\left(\frac{1}{2(\tau + T_D)}\right) + A_3^2 S\left(\frac{3}{2(\tau + T_D)}\right) \cdots\right]$$

or

$$S\left(\frac{1}{2(\tau + T_D)}\right) = (2N)(\tau + T_D)\left(\frac{1}{A_1}\right)^2 \sigma_H^2(\tau, T_D, N) - (A_3/A_1)^2 S\left(\frac{3}{2(\tau + T_D)}\right) - (A_5/A_1)^2 S\left(\frac{5}{2(\tau + T_D)}\right) \cdots . \tag{7}$$

This shows that some measurement of the noise at the third and higher harmonics must be made before noise at the fundamental can be calculated.

Acknowledgements

Much of this work was the outgrowth of a conversation with Dr. James Barnes of the National Bureau of Standards. The author is also indebted to Mr. Richard Ollins of the Hewlett-Packard Company for assistance in experimental work on the Hadamard variance.

References

1) Barnes, Chi, Cutler, Healey, Leeson, McGunigal, Mullen, Smith, Sydnor, Vessot, and Winkler, Characterization of Frequency Stability, NBS Technical Note 394, October 1970.

2) W. K. Pratt, J. Kane, H. C. Andrews, Hadamard Transform Image Coding, IEEE Proceedings, vol. 57, pp. 58-68, January 1969.

3) P. I. Richards, Computing Reliable Power Spectra, IEEE Spectrum, vol. 4, pp. 83-90, January 1967.

INSTANTANEOUS FREQUENCY

FIGURE 1

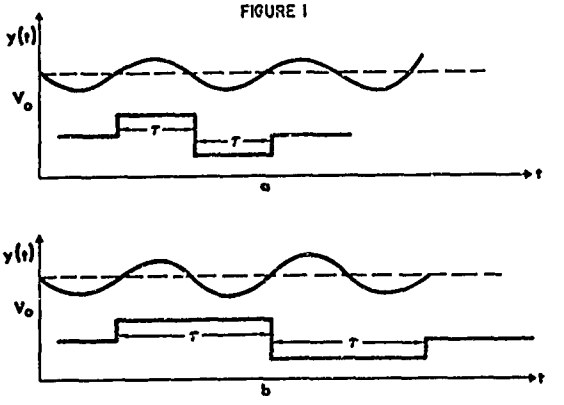
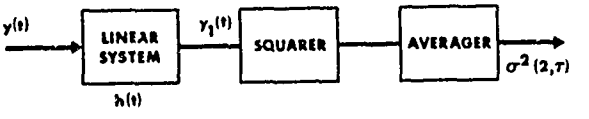


FIGURE 2



$H(f) = \mathcal{F}\{h(t)\}$, THE FOURIER TRANSFORM

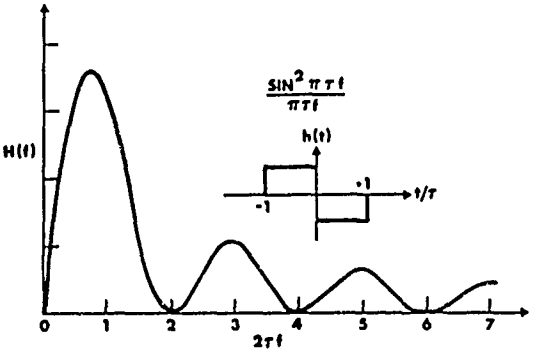
$y_1(t) = y(t) * h(t)$, CONVOLUTION

$S_{y_1}(f) = S_y(f)\,|H(f)|^2$, SPECTRAL DENSITY

$$\sigma^2 y = \int_0^\infty S_y(f)\,|H(f)|^2\,df$$

THE ALLAN VARIANCE TRANSFER FUNCTION

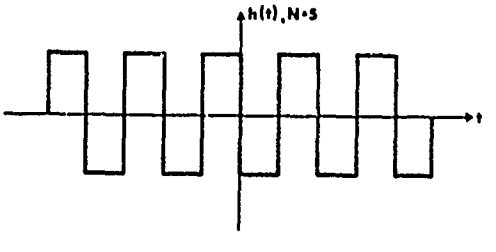
FIGURE 3



THE HADAMARD VARIANCE

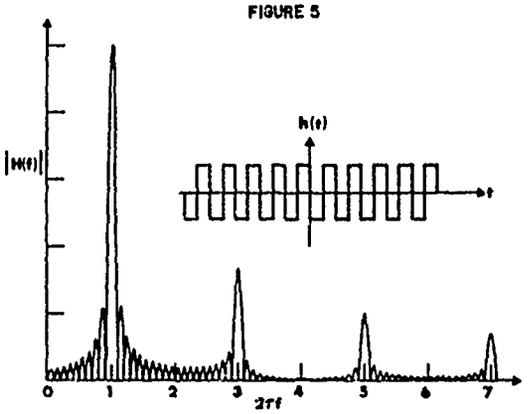
FIGURE 4

$$\sigma_H^2(N,\tau) = \left\langle\left[\bar{y}_{k+1} - \bar{y}_{k+2} + \bar{y}_{k+3} \cdots - \bar{y}_{k+2N}\right]^2\right\rangle$$



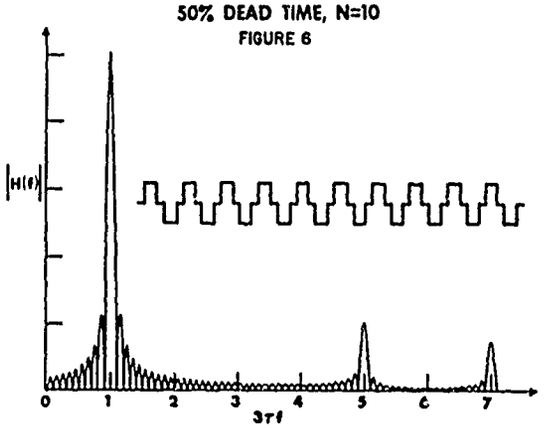
THE HADAMARD VARIANCE TRANSFER FUNCTION, N=10

FIGURE 5



THE MODIFIED HADAMARD VARIANCE
50% DEAD TIME, N=10

FIGURE 6



MODIFIED HADAMARD VARIANCE, BINOMIAL
COEFFICIENT ENVELOPE, N=5

FIGURE 7

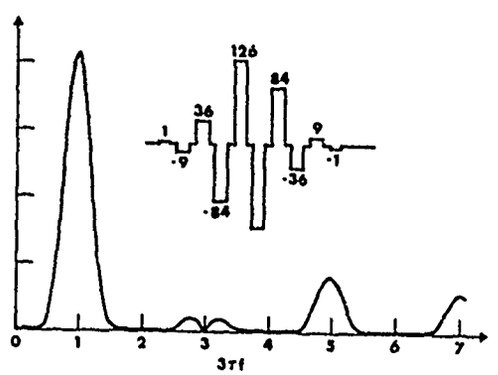
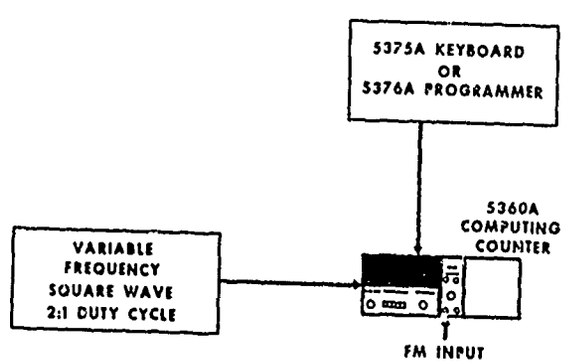


FIGURE 8



HIGH QUALITY QUARTZ CRYSTAL OSCILLATORS: FREQUENCY DOMAIN AND TIME DOMAIN STABILITY*

Helmut Brandenberger and Frederic Hadorn

Ebauches S. A.
Département Technique, Groupe Etalons de Fréquence
2001 Neuchâtel, Switzerland

and

Donald Halford and John H. Shoaf

Atomic Frequency and Time Standards Section
National Bureau of Standards
Boulder, Colorado 80302 USA

Summary

We measured the frequency stability of a pair of commercial 5-MHz quartz crystal oscillators which incorporate improved electronic design for enhanced short-term stability. The spectral density (frequency domain) of the phase noise, per oscillator, measured by each of our two laboratories, is

$$S_{\delta\phi} = + (10^{-11.8}\ \text{radians}^2\ \text{Hz}^0)\frac{1}{f^2} + (10^{-12.5}\ \text{radians}^2)\frac{1}{f} + (10^{-14.4}\ \text{radians}^2\ \text{Hz}^{-1})\, f^0,$$

over the range of about 10^{-1} Hz to 10^{+3} Hz.

Key Words (for information retrieval): Allan variance, Flicker of phase noise, Frequency stability, Oscillator noise models, Phase noise spectral density, Quartz crystal oscillator, Time domain stability.

Discussion

There are several aspects of electronic circuit design which must be carefully considered in order to achieve high short-term frequency stability in quartz crystal oscillators. We report an advance of more than ten decibels in the state-of-the-art for high quality ≈5 MHz quartz crystal oscillators in the noise frequency range of 1 to 100 Hz. The most important design factors were A) measurement and selection of transistors for the lowest possible flicker of phase noise[1] (oscillating loop and buffer stages), B) measurement and selection of transistors and diodes for the lowest possible DC flicker noise (voltage regulators, automatic gain control), C) massive negative feedback (DC and RF) in the RF circuitry to stabilize the RF gain and to reduce the flicker of phase noise of the transistors[1] (automatic gain control [AGC] amplifier, buffer amplifiers, oscillating loop amplifier). Because more negative feedback is used in the AGC amplifier and in the buffer amplifiers than can be used in the oscillating loop amplifier, the resultant flicker of phase noise performance of the oscillator is determined by the flicker of phase noise of the transistor in the oscillating loop.

Two commercial quartz crystal oscillators incorporating these design factors were measured in Neuchâtel, Switzerland and then shipped to Boulder, Colorado for further measurement. The spectral density (frequency domain) of the phase noise,[2,3] per oscillator, measured by each of our two laboratories, is

$$S_{\delta\phi} = + (10^{-11.8}\ \text{radians}^2\ \text{Hz}^2)\frac{1}{f^3} + (10^{-12.5}\ \text{radians}^2)\frac{1}{f} + (10^{-14.4}\ \text{radians}^2\ \text{Hz}^{-1})\, f^0, \quad (1)$$

over the range of about 10^{-2} Hz to 10^{+3} Hz. See Figure 1. From Figure 1 we see that there is gratifying agreement between our two laboratories in the measurements of the noise spectral density. The reproducibility of our frequency domain measurements is equal to or better than ±2 dB. Measurements made in the time domain (Allan variance[2,4]) gave results which were compatible with our frequency domain measurements. See Figure 2. This oscillator performance is obtained without the use of narrowband filters.

Each of our measurement systems (see Figs. 3, 4, 5, and 6) was based on the use of low noise double-balanced broadband mixers using Schottky-barrier diodes.[5,6] For most of the measurements, the oscillators are at zero-beat and in phase quadrature; the output of the mixer is amplified in a low-noise DC amplifier and sampled both in the frequency domain and in the time domain. The noise of our measurement systems can be measured easily; in each it is adequately lower than the noise of the oscillators.

The measured oscillator stability in the range 1 to 100 Hz is about 12 dB better than the prior state-of-the-art for oscillators of any type. The stability (time domain) in the 10 to 100 second range is not better than the performance of high quality quartz crystal oscillators which have been commercially available for the past eight years.[7,8] However, with a higher Q_L for the quartz crystal resonance, improved stability in the 10 to 100 second range may be expected.

*Contribution of the National Bureau of Standards, not subject to copyright.

To our knowledge, this has been the first successful application of low flicker of phase noise electronics to high quality quartz crystal oscillators. For examples of other devices, see references 1, 8, 9, and 10. Further improvement of the oscillator stability may be possible by further reduction of the multiplicative flicker of phase noise of the electronics. The present Q_ℓ is estimated to be about 1×10^6 on the basis of the noise measurements and is independently confirmed by an analysis of the circuit. A value of 3×10^6 is possible with 5 MHz crystals; this alone may allow a factor of three improvement in the 10 to 100 second stability, and may accrue in addition to the improvements in the electronics noise.

References

1. Donald Halford, A. E. Wainwright, and James A. Barnes, "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, and Cure," (Summary) Proc. 22nd Annual Symposium on Frequency Control, Fort Monmouth, N. J., April 1968, pp. 340-341.

2. J. A. Barnes et al., "Characterization of Frequency Stability," NBS Technical Note 394, October 1970; also published in IEEE Trans. on Instrumentation and Measurement IM-20, No. 2, pp. 105-120, May 1971.

3. L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," Proc. IEEE, Vol. 54, pp. 136-154, February 1966.

4. David W. Allan, "Statistics of Atomic Frequency Standards," Proc. IEEE, Vol. 54, No. 2, pp. 221-230, February 1966.

5. V. van Duzer, "Short-Term Stability Measurements," IEEE-NASA Symposium on Short-Term Frequency Stability, Washington, D.C.: U. S. Government Printing Office, pp. 269-272, NASA-SP80.

6. Donald G. Meyer, "A Test Set for Measurement of Phase Noise on High-Quality Signal Sources," IEEE Trans. on Instrumentation and Measurement IM-19, No. 4, pp. 215-227, November 1970 (Proc. of the 1970 Conference on Precision Electromagnetic Measurements, Boulder, Colorado, 2-5 June 1970).

7. Donald Halford, "Frequency Stability of Quality Quartz Crystal Oscillators: Performance and Some Critical Applications," Proc. of the Colloque International de Chronometrie, Paris, September 1969, Série A, pp. A-11-1 to A-11-3.

8. D. J. Glaze, "Improvements in Atomic Cesium Beam Frequency Standards at the National Bureau of Standards," IEEE Trans. on Instrumentation and Measurement IM-19, No. 3, pp. 156-160, August 1970.

9. D. G. Meyer, "An Ultra Low Noise Direct Frequency Synthesizer," Proc. 24th Annual Symposium on Frequency Control, Fort Monmouth, N.J., April 1970, pp. 209-232.

10. C. Finnie, R. Sydnor, and A. Sward, "Hydrogen Maser Frequency Standard," Proc. 25th Annual Symposium on Frequency Control, Fort Monmouth, N.J., April 1971 (to be published).

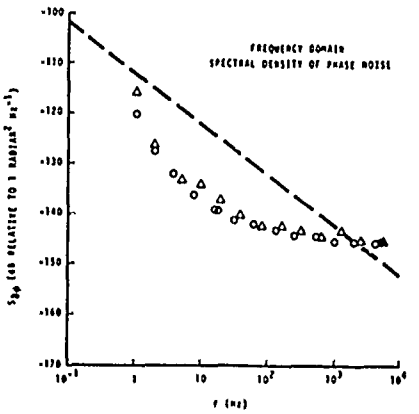


Figure 1 Spectral density of the phase noise of the improved commercial quartz crystal oscillator. The circles represent measurements made at Neuchâtel, Switzerland. The triangles represent measurements made at Boulder, Colorado. The measurements were made using 1 Hz bandwidth. The dashed line, in the region of 1 to 100 hertz, represents the prior state-of-the-art for quartz crystal oscillators.

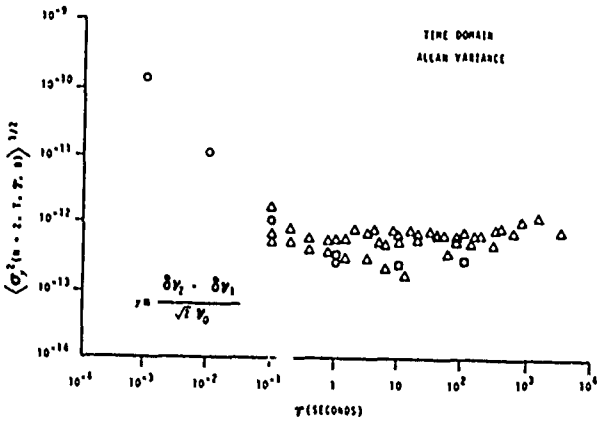


Figure 2 The square root of the Allan variance of the fractional frequency noise y is plotted against averaging time τ for the improved quartz crystal oscillator. The circles (zero-beat method) and the squares (non-zero-beat method) represent measurements made at Neuchâtel, Switzerland. The triangles represent zero-beat method measurements made at Boulder, Colorado. For all points, N = 2. For the circles, B ≈ 1000 Hz and the dead time is about 0.1 second except at τ = 1 second where the dead time is 1 second. For the squares, B ≈ 1000 Hz and the dead time is 1 second. For the triangles, B ≈ 25 Hz and the dead time is negligible.

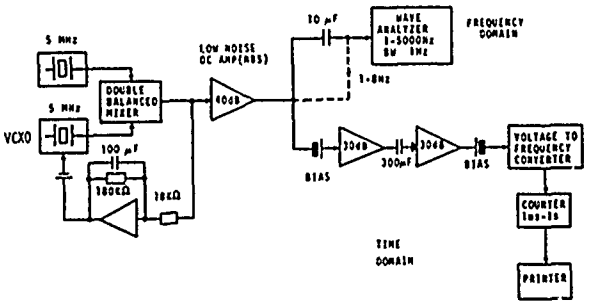


Figure 3 This setup was used at Ebauches for time domain measurements between 1 ms and 1 s (see Fig. 2). The voltage-controlled crystal oscillator (VCXO) is very loosely servoed (DC amplifier gain is about 10, servo loop time constant is about 18 s). The first amplifier following the mixer is a special design for low DC flicker noise. The voltage-to-frequency converter has a full-scale output pulse rate of 100 kHz.

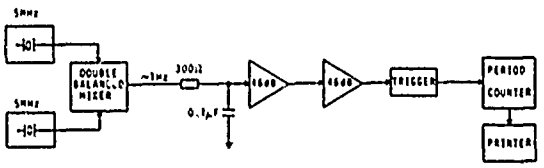


Figure 4 This setup was used at Ebauches for time domain measurements for sampling times τ of one second and greater (shown in Fig. 2). The trigger module is a Schmitt circuit of special design for low DC flicker noise.

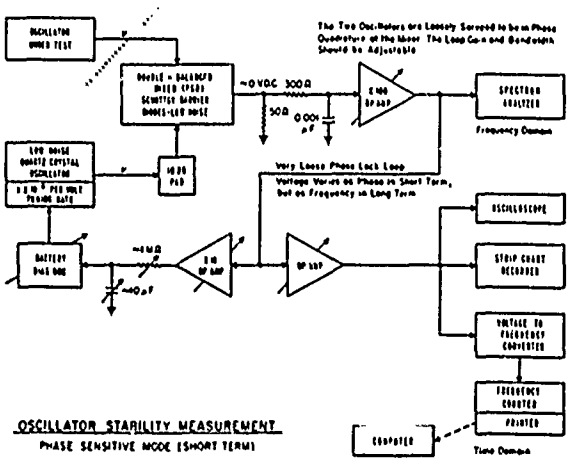


Figure 5 The frequency domain measurements by NBS (shown in Fig. 1) used this general method. The spectrum analyzer tunes from 1 to 5000 Hz and has measurement bandwidths of 1, 10, and 100 Hz. The first amplifier following the mixer is a special design for low DC flicker noise.

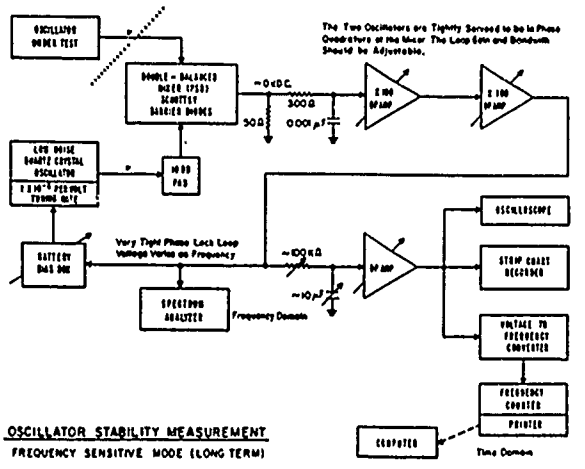


Figure 6 The time domain measurements by NBS (shown in Fig. 2) used this general method. The voltage-to-frequency converter has a full-scale output pulse rate of 100 kHz. The first amplifier following the mixer is a special design for low DC flicker noise.

THE DESIGN AND PERFORMANCE OF AN ULTRA-PURE, VHF FREQUENCY SYNTHESIZER FOR USE IN HF RECEIVERS

Max E. Peterson
Collins Radio Company
Dallas, Texas

Summary

This paper covers the design of a spectrally pure, VHF synthesizer for use in HF transceivers. The block diagram is discussed, showing how 100 Hz frequency steps are generated, and details of circuit design and construction are presented.

Introduction

We have known for some time that the best way to build HF receivers is to up-convert at the first mixer, doing all of the tuning with the first injection. The basic requirements for this scheme are

(1) a mixer with low intermod products,
(2) a synthesizer with low enough spurious signals and noise.

FET's have helped to reduce the mixer intermod products, and the subject of this paper is a synthesizer that makes the up-converting receiver a reality.

Advantages of up-converting are

(1) virtual elimination of all image responses,
(2) elimination of RF tuning,
(3) a cleaner frequency plan with fewer spurious frequencies being generated.

Let's review some of the requirements of a modern HF receiver that set the specifications for the first injection signal. Figure 1 shows a typical receiver input, a low level desired signal in the presence of a high level undesired signal. With no RF tuning, both of these signals hit the first mixer unattenuated. For a first IF frequency of 109.35 MHz, the injection to the first mixer will be 109.35 - 4.0 MHz, or 105.35 MHz at a level of approximately 2.5 volts RMS. Now if a spurious is present on this injection signal at 105.30 MHz, two identical and interfering IF signals will be generated. It is necessary to keep the undesired signal at least 10 db below the desired IF signal. This means all synthesizer spurious frequencies 50 kHz or more from the desired synthesizer output must be down:

$$20 \log \frac{2.5}{17 \times 10^{-6}} = 104 \text{ db}$$

(17 μv is the level of the synthesizer spurious that when mixed with a .35 v input would generate an output 10 db below the desired 50 μv output.)

Since residual noise can also mix the undesired signal into the IF passband, the noise level of the synthesizer must be down -139 db (104 db + 10 log $\frac{3000}{1}$) when measured in a typical audio bandwidth of 3 KHz. The inband noise level (300 Hz to 3.3 KHz from the desired synthesizer output) should be down over 40 db if a single sideband receiver noise figure of 40 db is met.

Since many HF radios are used for data handling, phase stability of the synthesizer is very important. This is in effect a measure of the close-in noise level, 20 Hz to 300 Hz from the carrier. There seems to be an infinite number of ways to specify phase jitter, but one I like and one that approximates the use of the signal source in a Kineplex data terminal states:

"The average phase of a signal, when measured over a 22 ms period, shall not vary more than 7° when any two adjacent 22 ms periods are compared."

An alternate method of measuring phase jitter will be discussed later.

Most HF equipment today requires 100 Hz channel spacing, and extensive tests at Collins have shown that to simulate a continuous tuning effect in a general-purpose receiver with a synthesized local oscillator, frequency steps of no more than 100 Hz can be used. In conventional communication equipment tuned by manually selecting each frequency digit, the time required for the synthesizer to change frequency is not critical. However, to simulate continuous tuning, the synthesizer must change channels in 25 milliseconds or less. Anything faster is good, but for switching times longer than 25 ms, undesirable tuning chirps develop.

We can now list the specifications for the synthesizer that make up-conversion practical.

Frequency Range	80 - 110 MHz
Frequency Steps	≤100 Hz
Tuning Time	≤25 ms
Output Level	0.5 v RMS
In-band Noise (300 - 3000 Hz)	≤-50 db
Phase Jitter	≤7° p-p
Noise Level 50 KHz Out	≤-139 db below desired output (3 KHz bandwidth)
Spurious Levels	
±10 KHz	≤-75 db below desired output
±20 KHz	≤-90 db below desired output

Synthesizer Description

System Description

The frequency mixing scheme of the 718U HF product line (now in production at Collins Radio) for which the

subject synthesizer was designed, is shown in Figure 2. This resulted from the combined effort of several design groups to make maximum use of the frequencies that would be required for the synthesizer. To generate 100 Hz steps we knew we needed 10 KHz and 9.9 KHz. The lowest possible reference frequency that could be used was 990 KHz, but this was not an acceptable VCXO frequency. A compromise was made at 9.9 MHz. This provided reasonable IF frequencies and IF injection frequencies, and we settled on the block diagram of Figure 3.

The synthesizer consists of two phase-locked loops, with loop 1 locked to a 9.9 KHz reference and loop 2 locked to 10.0 KHz. Mixers were used to reduce the maximum operating frequency of each variable divider to a frequency within the capability of available TTL logic circuits. Obviously, the VCO frequency of loop 1 is determined by the receiver operating frequency and is equal to:

$$f_{VCO\text{-}1} = 109.35 \text{ MHz} - f_{RX} \quad (1)$$

109.35 MHz is the first IF frequency and f_{RX} is the frequency to which the receiver is tuned. The selection of the operating frequency for the variable counter of loop 1 was a compromise between the maximum variable counter frequency of loop 2 and the maximum bandwidth of the half-octave filters that are between the antenna and the first mixer. We knew that the output of the loop 1 mixer would be present at some level as sidebands on the synthesizer output. The last half-octave filter has a 12 MHz bandwidth which covers 18 to 30 MHz, so to prevent a spurious problem, we operated above 12 MHz.

Unfortunately, as the operating frequency at the output of loop 1 mixer goes up, so does the output of loop 2 mixer. The lowest convenient division ratio that resulted for the loop 1 divider was 1400, for an operating frequency range of:

$$f_{max\text{-}1} = (F_{R1})(N_{max\text{-}1}) = (9900)(1499) = 14.8401 \text{ MHz} \quad (2)$$

$$f_{min\text{-}1} = (F_{R1})(N_{min\text{-}1}) = (9900)(1400) = 13.8600 \text{MHz} \quad (3)$$

The tuning range of the loop 2 VCO can now be determined knowing that the two least significant digits of the loop 1 divider represent the 100 Hz and 1 kHz digits of the operating frequency.

If the synthesizer frequency that corresponds to an operating frequency of 2 MHz is 107.35 MHz, and the loop 1 divider is set for 1400 (100 Hz and 1 KHz are set to zero), the frequency of VCO-2 is:

$$f_{VCO\text{-}2} = 107.35\text{-}13.860 = 93.490 \text{ MHz} \quad (4)$$

When the operating frequency is 29.9999 MHz the synthesizer frequency is 79.3501, and the frequency of VCO-2 is:

$$f_{VCO\text{-}2} = 79.3501\text{-}14.8401 = 64.51 \text{ MHz} \quad (5)$$

As shown in Figure 3, the loop 2 VCO is mixed with 99 MHz to reduce the operating frequency of the loop 2 counters to:

$$F_2\text{min} = 99.000\text{-}93.490 = 5.51 \text{ MHz} \quad (6)$$

$$F_2\text{max} = 99.000\text{-}64.510 = 34.49 \text{ MHz} \quad (7)$$

Several examples (see Table 1) should show how the two loops work together to vary the output in 100 Hz, 1 KHz, and 10 KHz to 10 MHz steps.

If we want 100 Hz or 1 KHz steps both N_1 and N_2 are advanced, but for 10 KHz, 100 KHz, 1 MHz and 10 MHz steps, only N_2 is varied. The common 4-bit, BCD code is used to program the counters N_1 and N_2. This is not a problem since the only digits in N_1 that are programmable correspond exactly to the two least significant digits of the operating frequency. Not so for N_2.

If an operating frequency of 0.000 MHz is selected the value of N_2 drops to 351. This is the value that must be added to the operating frequency so that the information that sets N_2 is correct. The three is added to the MHz digit, five added to the 100 KHz digit, and one is added to the 10 KHz digit. For an operating frequency of 21.3765 MHz, the value of N_2 can be calculated from:

$$N_2 = (99\text{-}[(109.35\text{-}f_{RX(MHz)}\text{-}N_1(.0099)])100 = 2553 \quad (8)$$

The breakdown of N_2 is as follows:

351 (Fixed Number)
+ 65 (100 Hz and 1 KHz Digits of f_{RX})
2137 (10 KHz-10 MHz Digits of f_{RX})
2553 TOTAL

There are two solutions to this problem of correlating the operating frequency to the division ratio of N_2. We can add the three numbers and program the counter with the sum of these numbers, or the counter can be programmed three separate times for each counting cycle. Because of the difficulty in adding BCD numbers and the complex circuitry required, we decided to reset the counters in N_2 three times for each 10 KHz output pulse. Fortunately, the minimum division ratio of the MHz- 10 MHz portion of N_2 is 3 (351), providing us with three different logic states (0, 1, 2), each of which is used to reset the counters in N_2 to a given state.

The block diagram of Figure 4 is typical of most digital phase-locked loops. The open loop bandwidth is given by:

$$BW = \frac{K_\phi K_V F_{(s)}}{N} \quad (9)$$

which is derived from the general transfer function of:

$$\frac{\Delta\phi_o}{\Delta\phi_i} = \frac{K_\phi K_V F_{(s)}}{s + \frac{K_\phi K_V F_{(s)}}{N}} \quad (10)$$

The amount of filtering of F_R by the filter $F_{(s)}$ for a sideband level in db (A) at the VCO output is

given by:

$$\text{attenuation } (F_s) = 20 \log \frac{K_\phi K_V}{F_R} + 6 \text{ db} + A_{(db)} \quad (11)$$

Finally, the frequency at the counter (N) input is always:

$$F_C = N \cdot F_R \quad (12)$$

For Figure 4, if N is increased by 1, the output frequency will increase by an amount equal to F_R.

Circuit Description

It is not my intent to show or discuss all of the circuitry that is used in this synthesizer. However, some of the circuits are unique and will be covered by this paper.

Phase Detectors. For certain phase-locked loop conditions, the phase detector of Figure 4 can be a simple set-reset flip-flop. However, as the tuning range of the VCO increases to beyond 10%, the phase detector must have the additional capability of frequency discrimination, because the capture range of a single phase-locked loop is less than its pull range. Various factors influence the capture range, such as:

(1) phase shift of loop filter,
(2) the value of K_ϕ, the transfer function of the phase detector,
(3) the value of K_V, the transfer function of the VCO,
(4) the range of N, the variable counter ratio.

As an example, the loop 2 of Figure 3 has a pull range of:

$$\Delta f_2 = 93.490 - 64.510 = 28.98 \text{ MHz} \quad (13)$$

The capture range is approximately 8 MHz, indicating that additional circuitry must be utilized for reliable operation.

This problem can be solved in two ways, either by coarse-tuning the VCO or by using a frequency detector. We chose the latter. The desired characteristics of this detector are that if the VCO is too low or too high in frequency, its output will lock either low or high until the VCO is pulled near its desired frequency; then it returns to being a phase detector. This function can be accomplished in varying degrees by using digital flip-flops. However, we wanted this function in a single package, which was possible if only gates were used. The circuit of Figure 5 was submitted to Fairchild for MSI and a 0.4" x 0.5" package resulted. In all, thirty-two gates are used. Three gates provide the basic set-reset flip-flop phase detector function, and five gates are used to shape each input pulse. Thirteen gates detect the condition of unequal input frequencies and lock the single output gate either high or low to force the VCO back to the correct frequency. Five gates form an out-of-lock detector that indicates a logic 1 if the loop is not phase-locked to the reference input.

Code Combiner. A second MSI function composed of sixteen gates is used in the loop 2 variable divider. This circuit (Figure 6) combines the three numbers that are used to reset each decade counter. One of the three four-bit words is selected by raising one of the three word selection inputs to a logic 1. If input 1 is raised to logic 1, the word A_1, B_1, C_1 and D_1 then appears at the output. Two of these circuits are required for the loop 2 divider.

Voltage-Controlled Oscillator (VCO). Ultimately, the signal-to-noise ratio of the VCO will limit the performance of the receiver. Any noise at all seems to be undesirable, but a S/N ratio of 100 db (measured 50 KHz from the synthesizer output in a 3-KHz bandwidth) seems an acceptable value although 139 db was desired.

To realize these noise figures, an FET oscillator circuit was considered a necessity. The circuit of Figure 7 is a modified Hartley, and is used both in loop 1 and loop 2. A solid-state tuning capacitor is used for tuning the oscillators, and both tune their respective ranges for an input of 1-8 volts. The DC amplifier, Q_2, has a gain of 1 and generates a negative image of the 1-8 volts for an equivalent capacitor voltage of 2 to 16 volts. CR_2, a Schottky barrier diode, limits the dynamic swing of the FET gate.

Frequency Multiplier. 99 MHz is required both as a receiver injection and for the mixer in loop 2. Since the reference frequency is 9.9 MHz, a 10 to 1 frequency multiplier was the obvious solution. In Figure 8, Q_1 is biased class B and acts as a doubler. $C_{1,2}$, and $L_{1,2}$ make up a bandpass filter that is tuned to 19.8 MHz. Q_2 is a linear amplifier that drives the step recovery diode CR_1. The crystal filter provides 40 db of attenuation to the 19.8 MHz sidebands on the 99 MHz output. Dual output amplifiers provide separate outputs, one at 600 mv and the second at 100 mv.

Loop Mixers. The two mixers of Figure 3 are almost identical, the big difference being the operating frequency. Loop 1 mixer output covers 13.86 MHz to 14.8401 MHz, and the output of the loop 2 mixer covers the range of 5.51 - 34.49 MHz. Loop 1 mixer has a very tough requirement on isolation, specifically the level of input 2 showing up on input 1. For a receiver frequency of 13.8902 MHz, the input to the loop 1 counter is 13.8897 MHz, which is the difference between the two mixer inputs VCO-1 and VCO-2. When the VCO-2 signal feeds through to the VCO-1 input, these two frequencies will mix again in the first mixer, generating a pseudo input frequency of 13.8897 MHz, which is only 500 Hz away from the desired frequency. This generates a distinct "birdie" that can be heard at the receiver even though the antenna input is grounded. Although there are several of these "birdies" in the frequency range of 13-14 MHz, they are acceptable if their relative signal levels are less than 1 μv. These levels were reached by adding the amplifier Q_1 in Figure 9, which provided an additional 40 db of isolation to the mixer.

The mixer itself is an RCA-CA3028A which is followed by an LC low-pass filter and a second CA3028A biased for linear amplification. Figures 10 and 11 show the construction of the two mixers. Point-to-point assembly, along with a copious use of shielded compartments, insures a maximum amount of isolation between inputs. For the loop 1 mixer, 100 mv on input 2 at 80 MHz will be attenuated by >90 db at input 1.

Frequency Standard. The frequency standard is a 9.9 MHz, temperature-compensated, crystal oscillator. Compensation is accomplished by a thermistor-resistor

network (Figure 12) that varies the capacitance of a solid-state capacitor CR3. The oscillator Q_1 drives an FET buffer amplifier Q_2, which drives a saturated amplifier Q_3. The output is a square wave that can drive one digital TTL gate. Figure 13 is a plot of:

(1) the uncompensated frequency drift as a function of temperature,
(2) frequency drift after compensation.

Each standard is compensated to better than ± 5 parts in 10^7.

Synthesizer Performance

The complete synthesizer is packaged on four printed circuit boards, 4.25" x 6.825". The modules that are mounted on each of the four-layer boards are "nested" so that the total volume is 71 cubic inches. Figures 14-17 are photographs of the four complete boards. The VCO board has both VCO modules, the two loop low-pass filters, and three voltage regulators. There is one negative regulator and a separate Collins thin-film regulator for the +10 volts required to power each VCO and filter driver. Separate positive regulators were necessary to prevent coupling between the two loops, coupling that results in the generation of 100 Hz sidebands. Each of the low-pass filters have in excess of 120 db attenuation at 9.9 KHz and 10.0 KHz. Pass-band ripple is 1 db and the 90° phase shift point is at 1080 Hz. This is an important frequency because it determines the maximum value of each loop bandwidth and it is important that the loop bandwidth be as high as possible. The purpose of a synthesizer is not to "clean up" a reference source, but to generate clean signals. A very narrow loop bandwidth, which isolates the synthesized output from the reference source, is redundant when the reference is many orders of magnitude cleaner than the synthesized output can ever be. A high loop bandwidth means faster capture time, lower in-band noise, and less susceptibility to magnetic radiation and vibration.

The mixer board has both loop mixers, and the variable counter for loop 1. The reference board contains the frequency multiplier, frequency standard, the dividers that provide 10 KHz, 9.9 KHz and 450 KHz outputs, and the two phase detectors. The divider board has integrated circuits only, which make up the loop 2 variable divider, N_2.

Noise Levels

Noise is usually measured at two points. We measure single-sideband inband noise in a 3-KHz bandwidth, 300 Hz to 3.3 KHz, and out-of-band noise at 50 KHz from the desired carrier frequency in a 3-KHz bandwidth. Figure 18 shows the set-up for measuring inband noise and Figure 19 is a plot of the data taken with an HP302 in 100 Hz steps. This integrates out to a value of 54.4 db. The noise 50 KHz out from the carrier is measured directly and equals -100 db.

Spurious Levels

Figure 20 is a plot of the spurious levels at an output frequency of 89.996 MHz. The expected 100-Hz sideband levels were <-62 db, 9.9 KHz sidebands were down -75 db, and 19.8 KHz sidebands were down <-90 db. Sideband levels beyond 19.8 KHz are too low to measure with this set-up. 60 Hz and 120 Hz sidebands, due to power supply ripple, were below the noise level of -62 db.

Phase Jitter

Phase jitter can be measured and specified in many different ways. One method that gives a quick, qualitative analysis of the phase jitter is shown in Figure 21.

The synthesizer and the HP5105 are phase-locked and set to the same frequency. The output of the mixer is a DC voltage with an AC voltage superimposed on it. It is the AC voltage that represents the phase jitter on the synthesizer output. Since a common integration period for phase jitter is 20 ms, the scope sweep speed can be set for 20 ms and a picture taken of one sweep. Figure 22 is such a picture. The vertical calibration is 4.5°/Cm. The average phase during this 20 ms is -0.19°, and the peak phase difference within this 22 ms period is only 3.3°. In a different test, readings of 1.5° were obtained using an analog test set that integrated the phase over 22 ms periods, and then displayed the difference between consecutive 22 ms periods.

Capture Time

The time it takes for a synthesizer to acquire phase lock on a new frequency is normally not a problem since most synthesized radios are controlled from manually switched control heads. However, one application for this synthesizer is in a scanning receiver. It was experimentally determined that 25 ms was the maximum time allowed if a "continuous tuning" effect was to be obtained. Figure 23 shows the test set-up for making these measurements. The receiver audio is first zero-beat, then the synthesizer is offset the desired amount (100 Hz, 1 KHz, etc.). As the switch is returned to its original position, the scope is triggered and the time is recorded for the receiver to reach zero beat again. The capture time for various frequency steps is listed below.

Frequency Steps	Capture Time
100 Hz	5 ms
100 KHz	10 ms
1 MHz	15 ms
10 MHz	25 ms

Conclusions

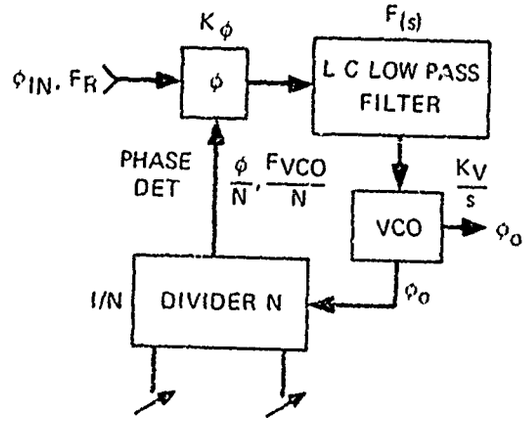
It is now possible to design synthesizers with the spectral purity necessary for use in up-converting HF receivers, and yet small enough to be practical. This synthesizer is now in production for use in an HF line of communication equipment that is considered to out-perform any other HF radio system previously built by Collins Radio Company.

F_{RCVR} MHz	F_{SYNTH} MHz	N_1	F_{N1} MHz	F_{VCO-2} MHz	N_2
2.0000	107.3500	1400	13.8600	93.490	551
2.0001	107.3499	1401	13.8659	93.480	552
2.0011	107.3489	1411	13.9689	93.380	562
2.0100	107.3400	1400	13.8600	93.480	552
2.1000	107.2500	1400	13.8600	93.390	561
0.0000	109.3500	1400	13.8600	95.490	351

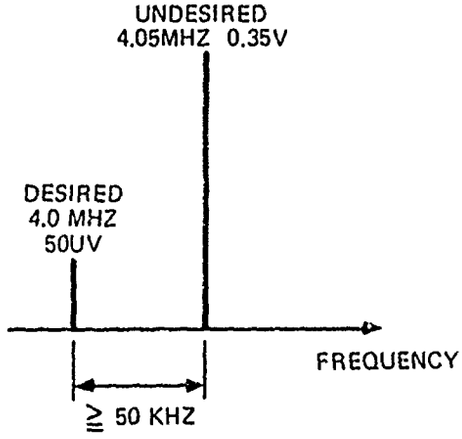
TABLE 1

GENERATION OF 100 HZ FREQUENCY STEPS

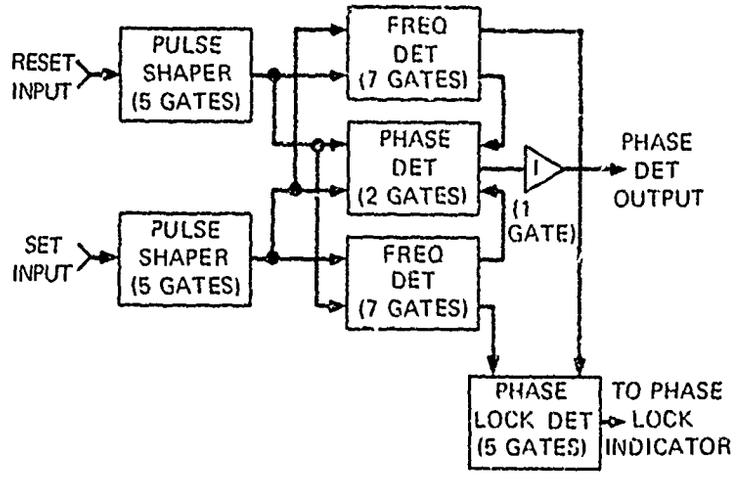
TYPICAL PHASE LOCKED LOOP



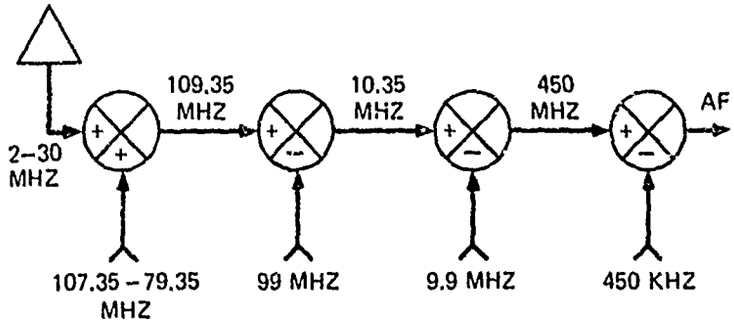
TYPICAL RECEIVER INPUT



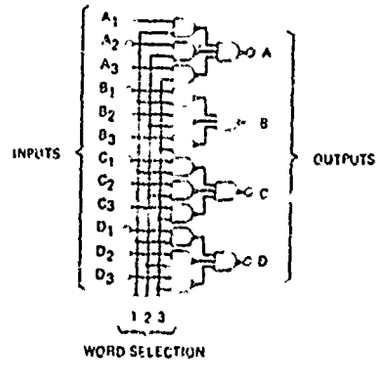
MSI PHASE DETECTOR



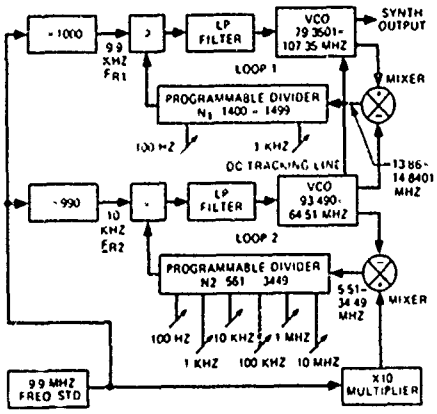
718U RECEIVER MIXING SCHEME



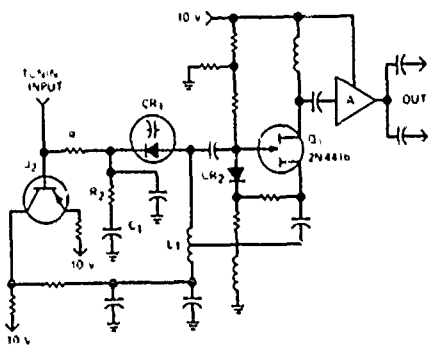
MSI CODE COMBINER

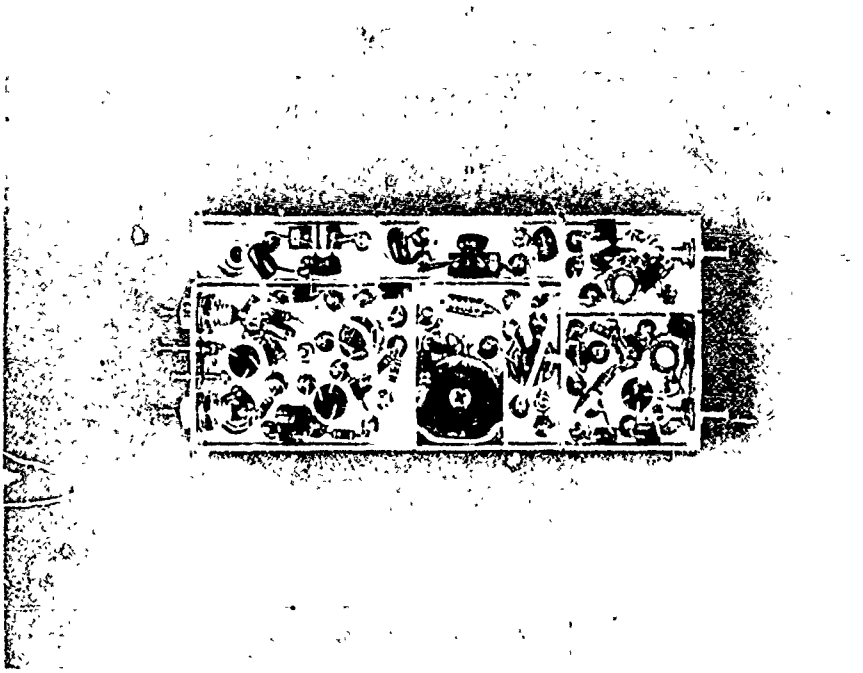


SYNTHESIZER BLOCK DIAGRAM

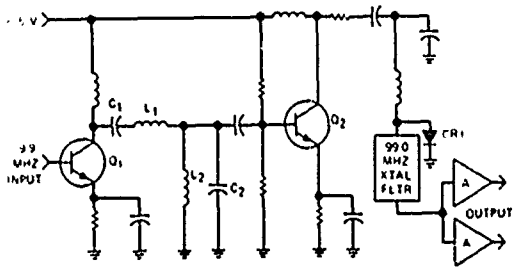


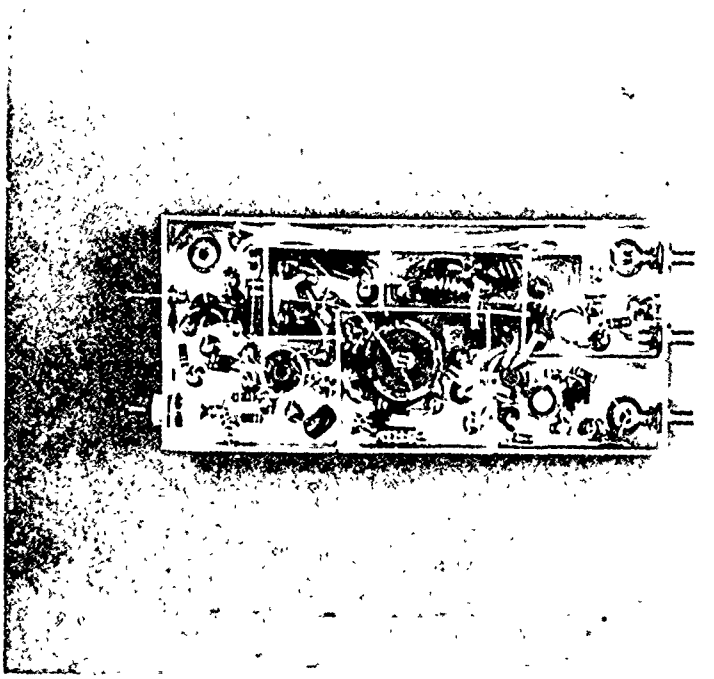
VCO CIRCUIT



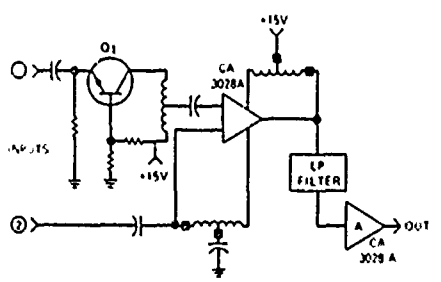


FREQUENCY MULTIPLIER

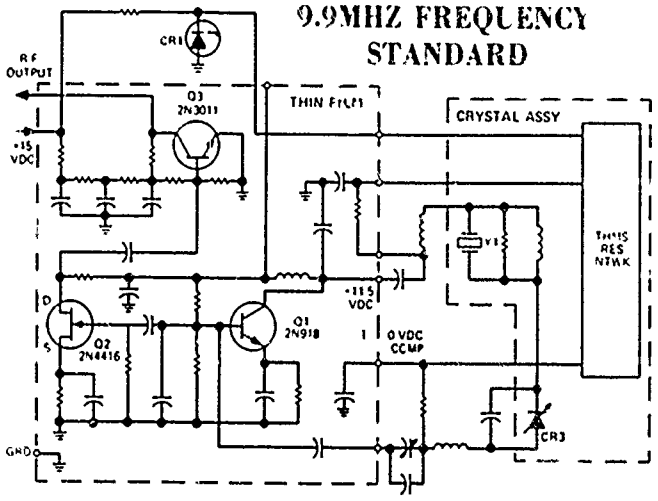




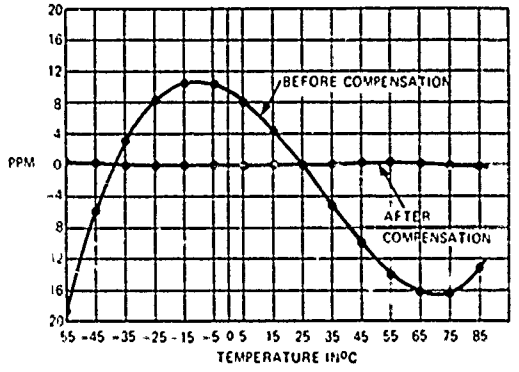
LOOP MIXERS

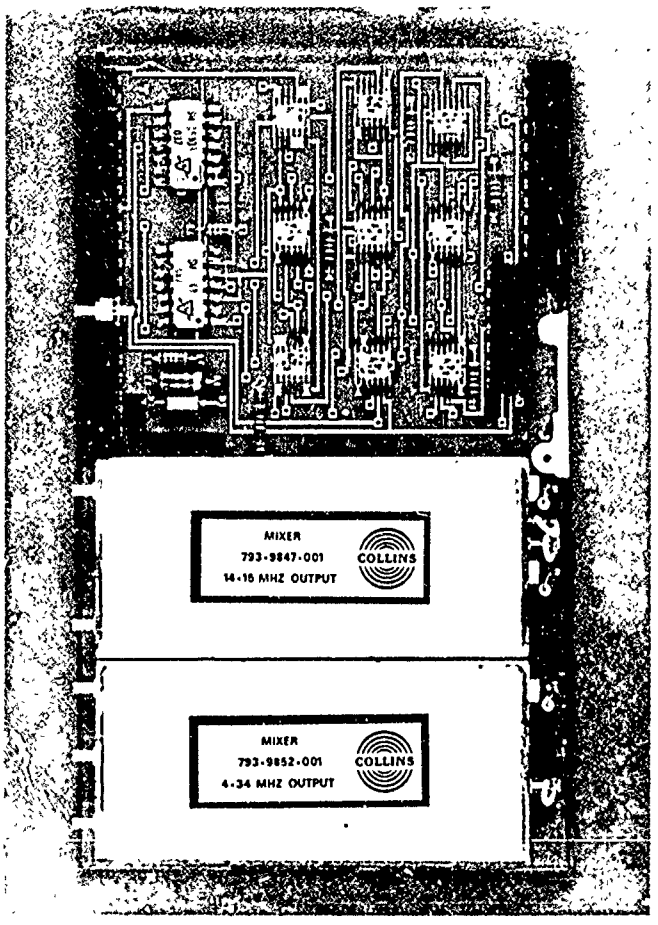


9.9MHZ FREQUENCY STANDARD

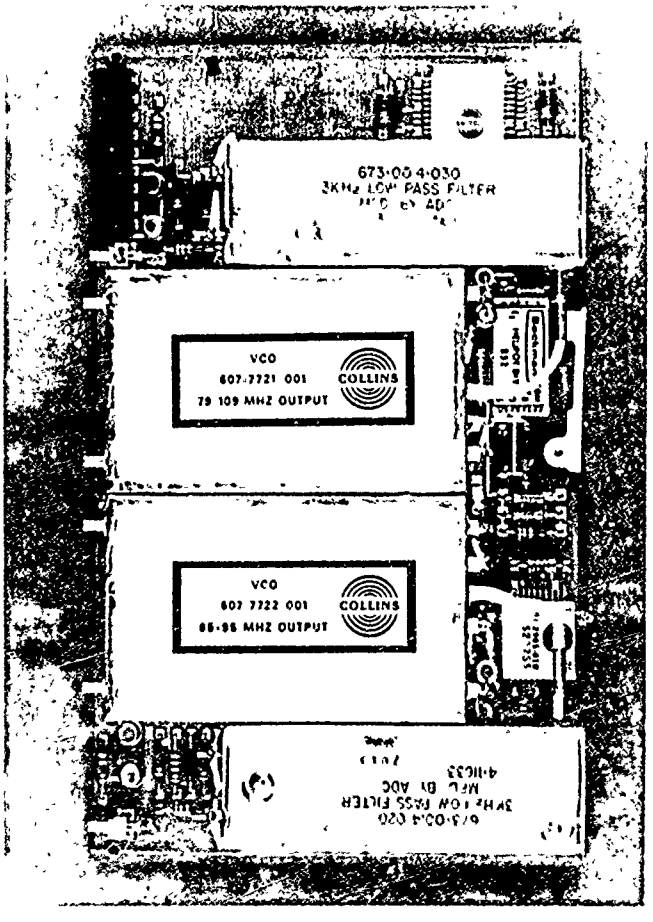


COMPARISON OF FREQ TEMP CHARACTERISTICS OF FREQ STD

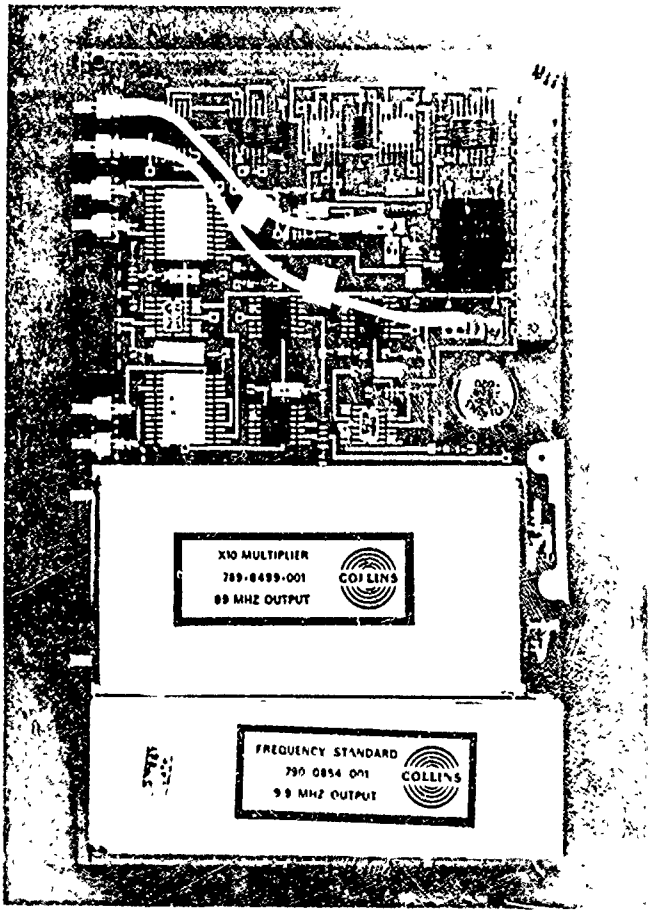


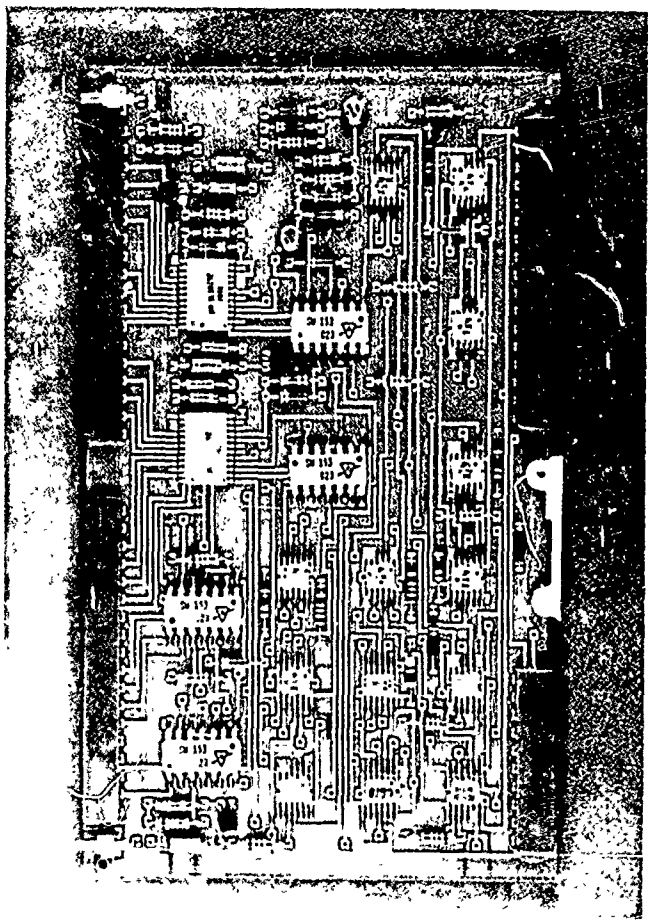


MIXER BOARD

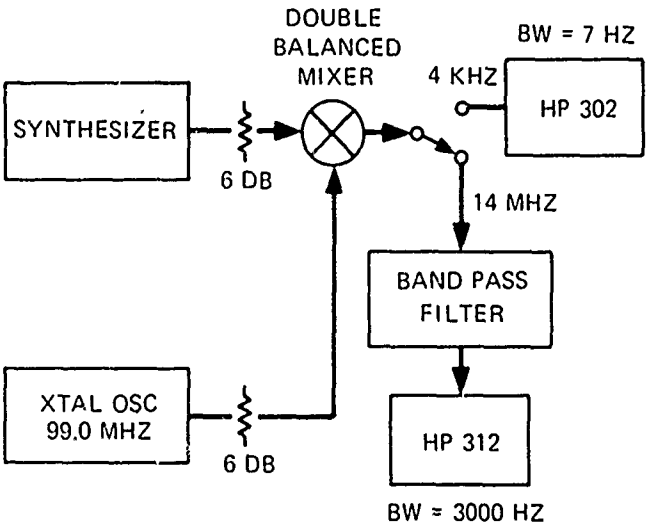


VCO OSCILLATOR BOARD

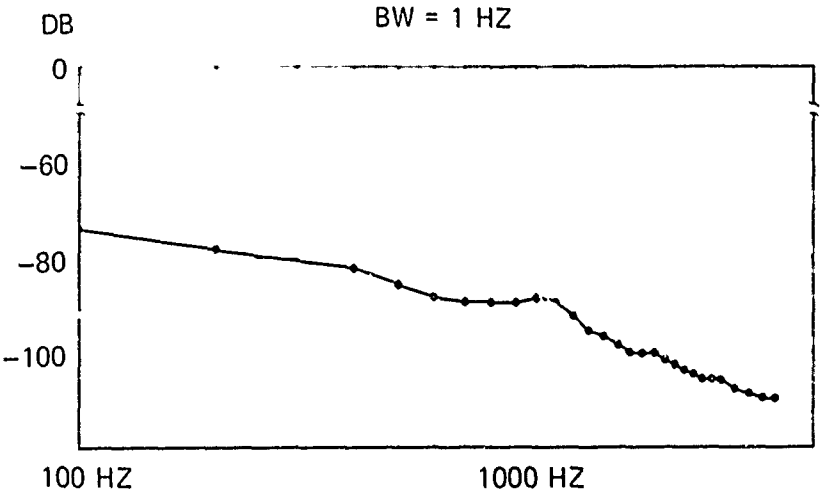




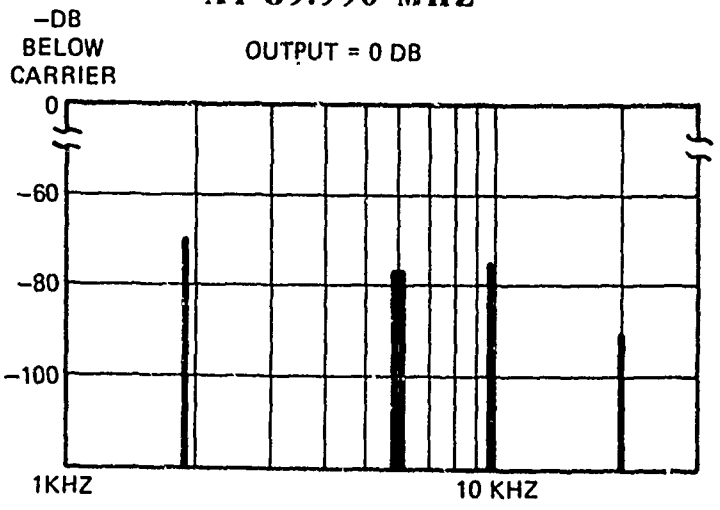
NOISE AND SPURIOUS TEST SETUP



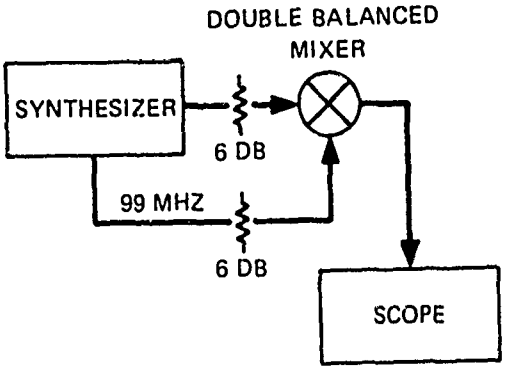
SYNTHESIZER NOISE DATA AT 89.996 MHZ



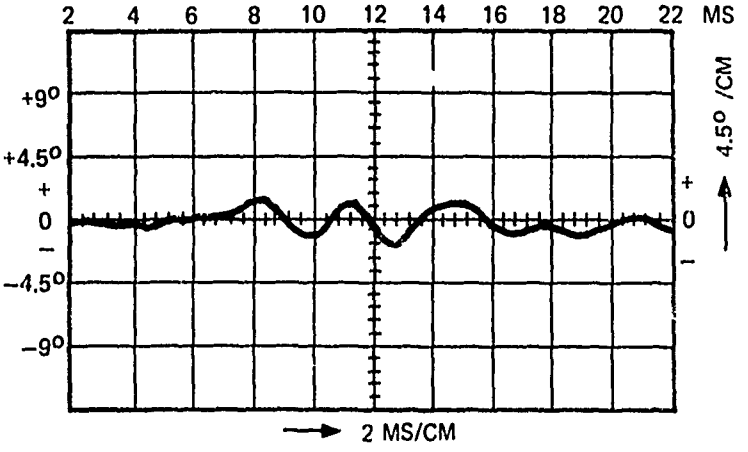
SYNTHESIZER SPURIOUS RESPONSE AT 89.996 MHZ



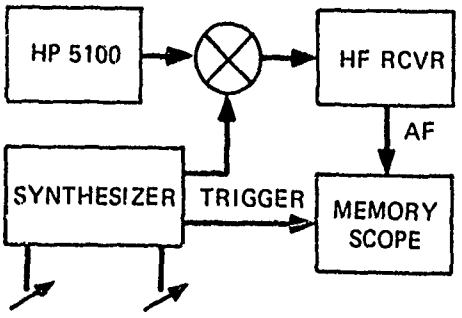
PHASE JITTER TEST SETUP



PHASE JITTER FOR ONE 22 MS PERIOD



CAPTURE TIME TEST SETUP



THE SPECTRAL FREQUENCY SYNTHESIZER

Dieter R. Lohrmann
Arthur R. Sills

Communications/ADP Laboratory
USAECOM, Fort Monmouth, N. J.

Summary

The synthesizer described in this paper resulted from the need for low power drain in battery operated portable radio sets. It is a digital type, yet requires only low speed digital circuitry even to the Gigacycle frequency range and higher. Hence, inexpensive MOS logic is used. The power consumption of the breadboard model is 100 mW. Power drain is reduced by eliminating the need for dividing the frequency to be controlled by a programmed counter and comparing the result with some reference frequency. Instead, the following principle is used: Starting from a known reference point, a voltage-controlled variable-frequency oscillator (VCVFO) is swept across a crystal-derived frequency line spectrum. The spectral lines have a spacing equal to the desired channel spacing and cover the frequency range desired. The number of spectral lines crossed during the tune-up sweep cycle are counted. When the desired count is reached, the sweep of the VCVFO is turned off, thereby locking it to this particular spectral line.

A model of such a synthesizer was built, covering the range from 41.5 to 64.5 MHz, and having a 50-kHz channel spacing. The power consumption was 40 mW, excluding VFO and buffer amplifier. The spurious responses ± 50 kHz off carrier were 85 dB below the carrier; all others were lower than 93 dB. The rms carrier phase jitter was smaller than 9 Hz in a 30-kHz bandwidth. The frequency acquisition time was below one second.

The breadboard was packaged, using thick-film technology with complementary MOSFET- integrated circuits in chip-and-wire technique. The packaged model measured 25.4 by 30 by 38 mm. A detailed description was given in ECOM Report Nr. 3265.

The drawback of this first model was that it required a maximum of one second of frequency acquisition time which was considered too long. Further, some narrow tolerances in the VFO and tuning voltage dividers as required, were undesirable from the standpoint of cost in mass production and also in regard to reliability.

In our latest model we managed to avoid these problems on the penalty of increased complexity of the logic. However, this is not considered a drawback, since the cost per flip flop on a MOS LSIC is very low. This latest model which is in the breadboard stage now, first uses a 1 MHz spaced spectrum for the MHz line count, then switches to a 100 kHz spaced spectrum for the 100 kHz line count and finally turns on the 25 kHz spectrum for the 25 kHz count. Coarse and fine tuning voltages for the VFO are derived from the output of the sampling bridge. There are no resistor matrices or digital-analog converters required for deriving the coarse tuning voltage.

The description of this improved model, its principle of function and detailed description, using block diagram and schematics will then be presented.

Introduction

Many frequency synthesizers for radio equipment in the field today now use the classical ÷ N principle. However, for applications in the VHF frequency range and above the divide by N IC counters involved still have pretty high battery power consumption, such that the complete synthesizer requires in the order of watts of power drain. This high power consumption is not desirable in battery powered radios. Ideally one would like to get away with 500 mW of battery power drain or less for the entire receiver. Therefore, we came up with a scheme which would permit us to avoid high speed IC logic. In other words, we eliminated the now already classical divide-by-N-counter and associated circuitry.

The Principle

Shortly, the scheme works as follows:

A Multiline Frequency Spectrum is created which covers the frequency range desired, the spectral lines having a spacing which is equal to the channel spacing desired, for instance, for the range, e.g., from 40 to 67 MHz, with channel spacing of 25 kHz within that range, see Figure 1.

A variable frequency oscillator will then be locked to any particular one of these lines by phase lock techniques. Since the spectrum is crystal derived, effectively the VFO output frequency will be single crystal controlled. Figure 2 shows part of the basic arrangement. The output from the 4 MHz crystal oscillator is divided down to 25 kHz. The pulse generator is triggered by the 25 kHz signal, delivering a pulse with 25 kHz repetition rate and 7 ns pulse width. This pulse has a spectrum which contains harmonics of 25 kHz up to 67 MHz and higher. By means of the phase lock loop, which is comprised of the sample and hold circuit, the dc amplifier and the VFO, the VFO can be permanently locked to any of these spectral lines. This principle is well known and used in practice; sometimes it is called the Impulse Governed Oscillator, or for short, IGO. The problem which we solved is, to provide relatively simple, inexpensive and low power consuming means to lock the IGO to the particular spectral line desired. This is accomplished in the following manner:

In the beginning of the tuning procedure, the oscillator is preset to a known reference point in the spectrum. Starting out from this point, the oscillator is swept across this spectrum, counting the spectral lines crossed. Assume this reference be 40 MHz, then after e.g., 7 counts we should arrive

at 40 plus 7 X 25 kHz = 40.175 MHz. If this is the frequency that was desired, the sweep is turned off and the oscillator then locks to the 40.175 MHz spectral line. In the original model which was described in ECOM Report #3265, we had used a reference line every integer MHz, obtained by coarse presetting the oscillator by a resistor matrix in 1 MHz steps and using a 1 MHz spaced spectrum at the startout of the tuning procedure. This scheme had the disadvantage of requiring a frequency acquisition time of maximum 1 second, furthermore it required some precision in the preset circuitry which was considered objectionable in respect to cost and reliability.

These disadvantages were eliminated in the second generation model presently being discussed. Fig. 3 shows the block diagram of this model.

The tuning cycle is initiated by the "Start Tuning" switch whenever the frequency control knobs are operated. In the beginning of the tuning cycle, the 4 MHz crystal reference oscillator thru gates 1 and 5 triggers the pulse generator. Thus the spectrum of this pulse sequence contains harmonics on 36, 40, 44, 48, 52....MHz. As a reference point in the spectrum is chosen 40 MHz. Therefore, the logic, via the preset line presets the VFO to 39 MHz. Then the logic turns on the sweep generator which sweeps the frequency of the VFO upwards.

As the frequency of the VFO approaches the 40 MHz spectral line, the beat frequency at the output of the sampler decreases. This is sensed by the arrangement of high pass filter & detector #1; when the beat frequency has decreased to 80 kHz, Schmitt trigger #1 fires, whereupon the logic reduces the sweep speed of the VFO. As the frequency of the VFO gets still closer to the 40 MHz line, (about 12 kHz) Schmitt trigger #2 triggers, whereupon the logic reduces the sweep speed further.

As the VFO frequency hits the 40 MHz reference line, lock occurs. If 40.000 MHz was the frequency desired, the sweep is turned off. If not, then the information, how many 1 MHz lines have to be counted which is presented on the X 1 MHz input lines is stored into the memory of the logic. Then the logic closes gate 1 and opens gate 2, providing pulses with 1 MHz repetition rate to the sample & hold bridge, thus creating a 1 MHz spaced spectrum. Thereafter the logic turns on the sweeps again. Whenever an integer MHz spectral line is crossed, Schmitt trigger #1 provides a pulse to the line counter in the logic. When the count corresponding to the data on the X 1 MHz input lines is reached, e.g., 7, then the logic initiates the same locking procedure that was used for locking to the reference, hence lock to 47.000 MHz occurs. If this was the frequency desired, then after lock the logic turns off the lock sweep. If not, the X 100 kHz information from the input lines is stored into the logic, the spectrum is switched to 100 kHz spacing and the sweeps are turned on again, except for the fast sweep speed. Similarly, 100 kHz lines crossed are counted till the desired line is reached and lock occurs to that line. In like fashion, the 25 kHz information is processed.

The three sweep speeds used are: 1 MHz/ms for the 1 MHz line count, which is much slower than the theoretical limit, 50 kHz/ms for the 100 kHz line counting and 3 kHz/ms for the X 25 kHz count and lock.

Details of the Circuits

The logic is built with CMOS IC's. It contains 17 flip flops and 88 gates. It lends itself to construction in LSI MOS or CMOS technique.

The Pulse Generator (See Fig. 4)

It is really a two stage differentiating amplifier; the use of a step recovery diode under the restriction of low power consumption does not simplify the pulse generator.

The Sampling Bridge

It is a HPA 2356 Quad. The hold circuit consists of a 12 pF capacitor and a 2N 5566 dual balanced FET transistor, see Fig. 5. Use of the latter eliminates bias shifts of the tuning amplifier due to temperature change. A CA3060 is used as the DC Loop Amplifier. The first loop filter is comprised of R14, R15 + R16 and C19 + C20. The voltage present at this capacitor is swept by the sweep generator during the tuning procedure. Integrated Circuit NH0001, which interconnects the sweep generator with this loop filter capacitor is used to disconnect the loop from the sawtooth generator, releasing the loop after lockup.

The Sweep Generator

It consists mainly of capacitor C8 + C9 + C10 to which are fed three current levels by the logic, thereby creating three sweep speeds.

The Beat Note Amplifier

It could be simplified, provided an appropriate IC becomes available.

The Non Linear Tune Amplifier

It provides the tuning voltage for the varicaps in the VCO, which ranges from -12 to -42V, see Fig. 6. The counterfeedback of this amplifier is non-linear in order to get an approximately linear relationship between tuning voltage and output frequency of the VFO. This is done in order to keep the loop gain of the phase lock loop approximately constant over the frequency range of the oscillator.

A Second Loop Filter section is inserted between the control line of the VFO and the output of this nonlinear amplifier. This is required in order to reduce the inband noise of the VFO.

The Phase Lock Loop

The VFO is completely controlled by the output of the sampling bridge; there is no requirement for coarse tuning.

Fig. 7 shows the beat sensors which are high pass filters, detectors and schmitt triggers #1, 2 and 3.

Summing up, this frequency synthesizer has these particular features (see Figure 8):

Mostly digital design with inexpensive low speed logic suited for LSI.

Frequency range 40 to 67 MHz, suited for a transceiver for 30 to 80 MHz.

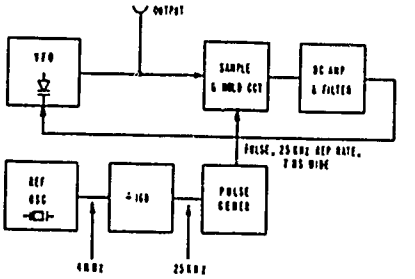
25 kHz channel spacing.

120 mW of dc power consumption, including VFO, buffers, and tuning voltage supply.

All digital controlled.

Fast frequency acquisition time (< 200 ms).

Single crystal controlled.



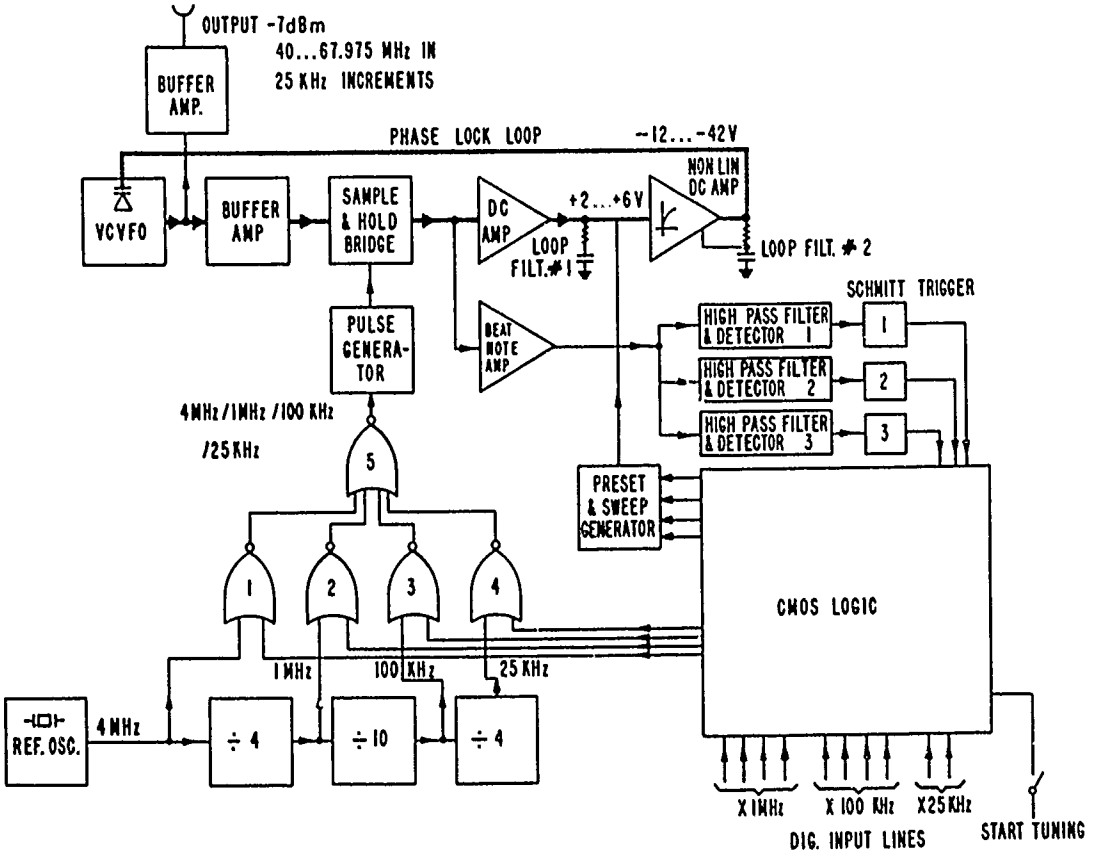


Fig. 3: BLOCK DIAGRAM OF SPECTRAL SYNTHESIZER

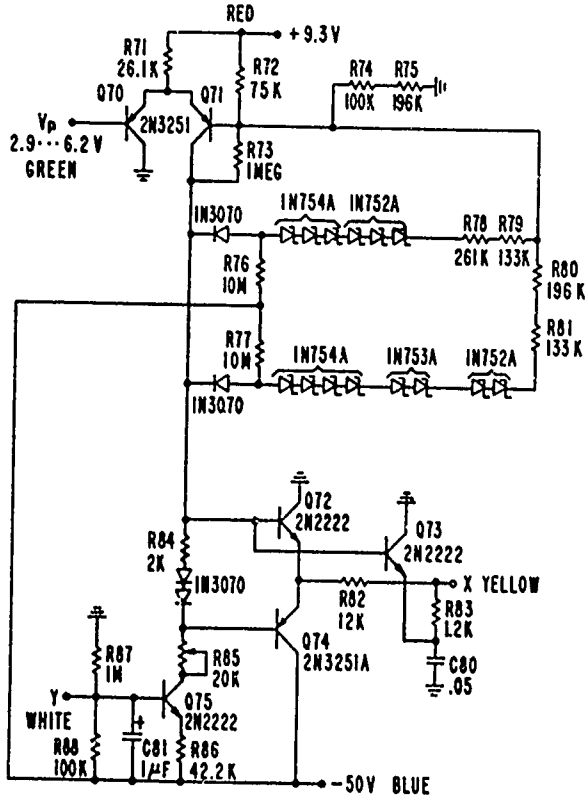


Fig. 6: NON LINEAR TUNE AMP

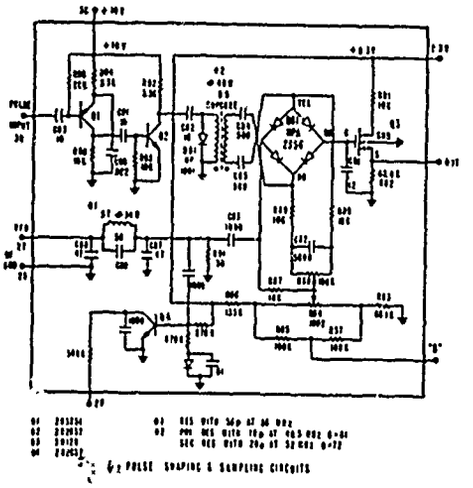


Fig 4 PULSE SHAPING & SAMPLING CIRCUITS

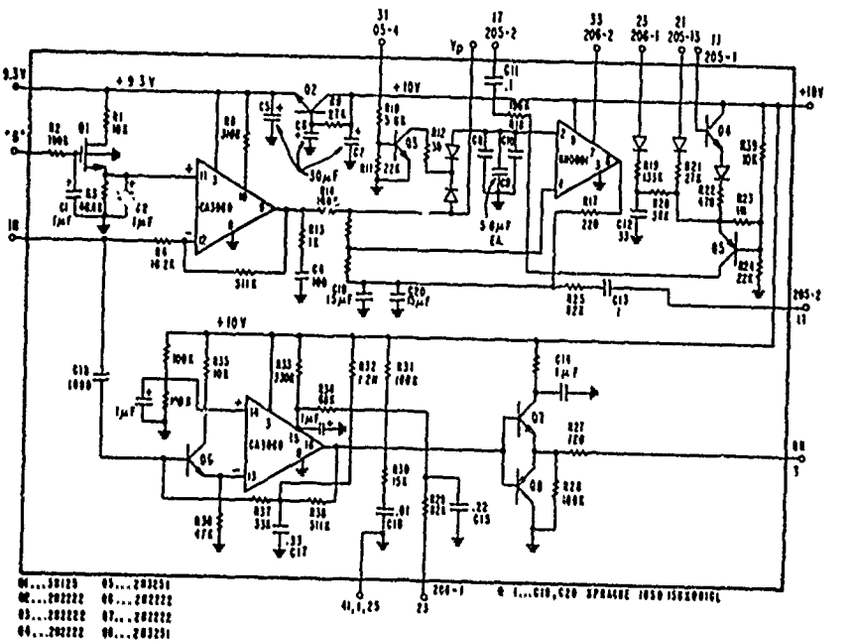


Fig 5 LOOP AMPLIFIER, BEAT NOTE AMPLIFIER & SWEEP GENERATOR

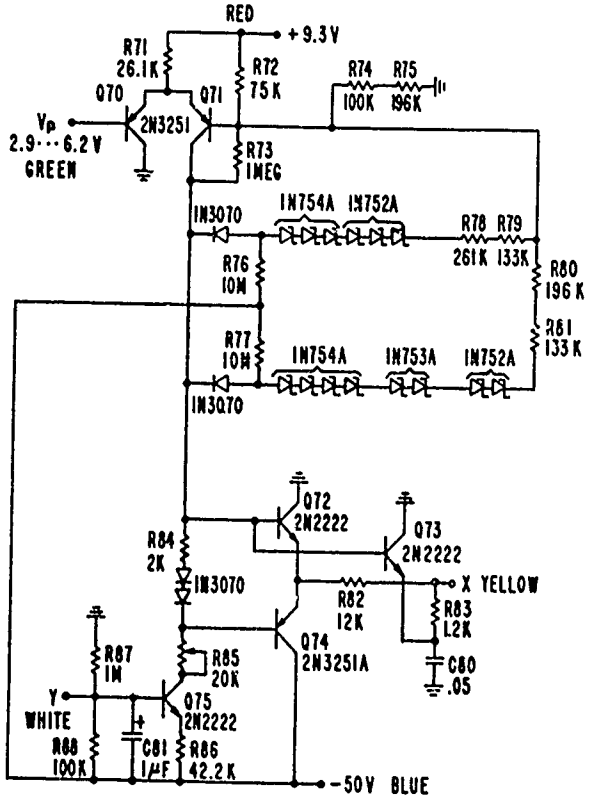


Fig. 6: NON LINEAR TUNE AMP

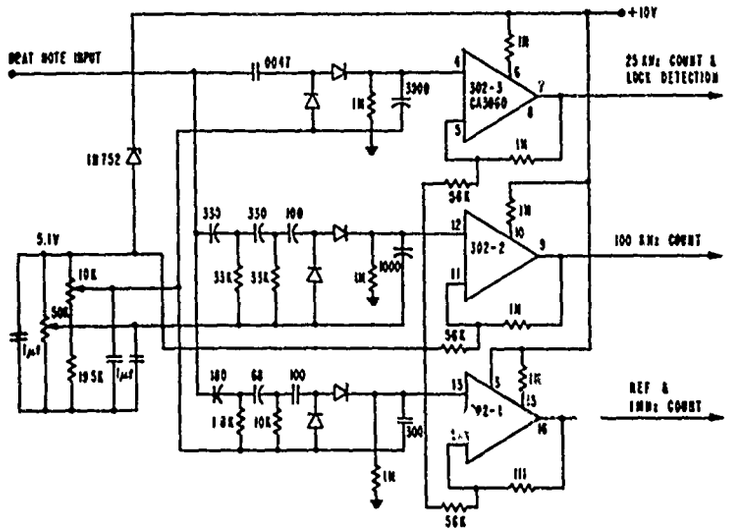


Fig. 7 : BEAT NOTE HIGH PASS FILTERS, DETECTORS AND SCHMITT TRIGGERS 1....3

MOSTLY DIGITAL DESIGN WITH INEXPENSIVE LOW SPEED LOGIC SUITED FOR LSI.

FREQUENCY RANGE 40 TO 67 MHz. SUITED FOR A TRANSCEIVER FOR 30 TO 80 MHz.

25 kHz CHANNEL SPACING.

120 mW OF DC POWER CONSUMPTION, INCLUDING VFO, BUFFERS AND TUNING VOLTAGE SUPPLY.

ALL DIGITAL CONTROLLED.

FAST FREQUENCY ACQUISITION TIME t < 200 ms).

SINGLE CRYSTAL CONTROLLED.

ENERGY TRAPPING IN A LITHIUM TANTALATE X-CUT RESONATOR

Ken'ichi Sawamoto

Musashino Electrical Communication Laboratory

Nippon Telegraph and Telephone Public Corporation

3 Tyome, Midoritye, Musasino-shi, Tokyo, Japan

Abstruct

Energy trapping effect is studied experimentally in lithium tantalate X-cut resonators. Resonant frequency lowering is mainly contributed by a piezoelectric effect, inversely by a mass loading effect in a quartz crystal.

Considering a suitable anisotropic constant, resonant frequencies of thickness twist- and thickness shear- mode are expressed by a similar formula as a function of the ratio a/H, where a is the length of electrodes and H the thickness of resonator.

In a monolithic filter, obtainable maximum band width is about 6 % , and corresponding a/H is 1.85 in case of thickness shear mode, and 1.04 in case of thickness twist mode.

Coupling factor depends on not only propagating mode but also the length a and width b of electrode.

Introduction

Recently monolithic crystal filters using energy trapping effect are being developed. These filters are hither-to made by quartz crystal. Because of its small coupling factor, these filters are usable only as narrow band pass filters. Lithium tantalate single crystal is thought as a suitable material to make a wide band pass filter with band width ratio larger than 1 % . An X-cut $LiTaO_3$ resonator can be regarded to excite only one pure shear mode, and its coupling factor is about 4 times larger than quartz. Furthermore, a resonator having a zero temperature coefficient at room temperature was obtained by adjusting the effective coupling factor to a specified value[1].

In order to design a monolithic filter, it is necessary to know the vibration energy distribution associated with an electroded area excited at resonance. In this paper, energy trapping were studied experimentally, and it is shown that the theory used in quartz resonator is applicable to case of an X-cut $LiTaO_3$ resonator. Also band width of monolithic filter was estimated.

The property of an X-cut $LiTaO_3$ plate

AH X-cut resonator belonging to the point group 3m has two shear modes. One is fast wave with velocity v_+ and coupling factor k_f, another slow shear wave with velocity v_- and k_s, respectively. These are expressed as follow,

$$\rho v_{\pm}^2 = (1/2)\left[(c_{66} + c_{44}) \pm \sqrt{(c_{66} - c_{44})^2 - 4 c_{56}^2}\right], \quad (1)$$

$$\frac{\tan\left(\frac{\pi}{2}\frac{f}{f_{of}}\right)}{\left(\frac{\pi}{2}\frac{f}{f_{of}}\right)} k_f^2 + \frac{\tan\left(\frac{\pi}{2}\frac{f}{f_{os}}\right)}{\left(\frac{\pi}{2}\frac{f}{f_{os}}\right)} k_s^2 = 1, \quad (2)$$

where ρ is density of the crystal, f resonant frequencies, f_{of} and f_{os} are anti-resonant frequencies corresponding to v_+ and v_-, respectively.

The directions of displacement corresponding to the velocities v_- incline at an angle of $\theta_{\mp}$ with y-axis, and are given by

$$\tan\theta_{\mp} = (v_{\mp}^2 - c_{66})/c_{56} . \quad (3)$$

The angle of inclination θ_+ was observed by T.Uno[5] using the light diffraction phenomenon due to an acoustic wave, and it is confirmed that $\theta_+ = -49^\circ \pm 2^\circ$. These physical constants for $LiTaO_3$ are given in Table 1.

The coupling factor k_f is about 9 times larger than k_s. Since acoustic power is proportional to the square of coupling factor, the slow shear wave can be neglected.

Experimental procedure

In an X-cut resonator, there are two types of wave, one is propagating to a direction of displacement and is named as thickness shear mode (TS mode), another wave propagates to a direction making an angle 90° with that of TS mode and named as thickness twist mode (TT mode). Therefore two types of specimens were used in this experiment. The electrode (evaporated Ni) configuration on the crystal plate is shown in Fig. 1, and the dimensions are given in Table 2.

In specimen A (Fig. 1), TS mode is propagated, and another specimen B was made to propagate TT mode. From the measurement of vector admittance diagram of specimen A, the mechanical quality factor Q, two terminal equivalent circuit L_1, C_1, R, and C_0 were obtained (Table 2), and the coupling factor of these resonators are estimated from the overtone frequencies.

The electrodes were divided into two equal

pairs along the dotted line (75 or 100 micron) as shown in Fig. 1 by YAG laser machining apparatus, and two resonators were constructed by thus the divided two pairs of electrodes. Since these two resonators were coupled in each other, resonant frequencies of two types were measured. One is a symmetric anharmonic thickness shear mode, and another an anti-symmetric mode.

After the measurements of resonant frequencies were completed, some portion of the end sides of the electrodes were cut by YAG laser and the separated Ni electrodes were dissolved by $FeCl_3$ solvent, so that the length a of the electrodes become short. The similar treatments were repeated 5 times on the specimen A and 4 times on the specimen B. Table 3 shows the length a of the electrode, the ratio of the a to the thickness H, and the resonant frequencies.

There are two cut off frequencies f_o and f_e, and f_o corresponds to the antiresonant frequency and is equal to $v_s/(2H)$, and f_e to the resonant frequency in a fully electroded plate. Now a resonant frequency lowering Δ is defined as

$$\Delta = (f_o - f_e)/f_o = P + R \quad , \quad (4)$$

where, P corresponds to a piezoelectric effect and calculated from Eq. 2 neglecting the term of slow wave, and R is mass loading effect expressed as $R = 2\rho' h/(\rho H)$, where ρ' and h are density and thickness of the evaporated metal electrode. It is impossible to fulfil the condition that a sufficiently wide area of electrode compared with thickness of the plate, then f_e is estimated from the lowest measured resonant frequency. Therefore, Δ obtained from TS mode differs from that of TT mode. Table 4 shows measured k, and R with calculated P. Although, frequency lowering Δ is somewhat larger than P + R, it is mainly contributed by the piezoelectric effect.

Discussion

When the coordinate is rotated around x_1 axis about θ_+, the axes of this plate correspond to that of the rotated y-cut quartz plate as follow

Direction	rotated Y-cut quartz	X-cut $LiTaO_3$
thickness	x_2 or 2	x_1 or 1
displacement	x_1 or 1	x_2 or 2

Therefore, the problems of the wave propagation for the plate is treated by similar methods as be done for the rotated Y-cut quartz plate[2].

In case of TT mode such as specimen B, a resonant frequency of anharmonic symmetric mode f is given by

$$(\Gamma_e/\Gamma_s) \tan (\Gamma_e a) = 1 \quad , \quad (5)$$

and for an anti-symmetric mode

$$(\Gamma_e/\Gamma_s) \cot (\Gamma_e a) = -1 \quad , \quad (6)$$

where

$$\Gamma_e = \left(\frac{m\pi}{\gamma_{TT} H}\right) \cdot \sqrt{(f/mf_e)^2 - 1} \quad , \quad (7)$$

$$\Gamma_s = \left(\frac{m\pi}{\gamma_{TT} H}\right) \cdot \sqrt{1 - (f/mf_o)^2} \quad , \quad (8)$$

and γ_{TT} is an anisotropic constant equal to $\sqrt{c'_{44}/c'_{66}} \simeq 0.8$, m is order of the mode. In this treatment, $c'_{66} = 1.42$, $c'_{44} = 0.95$, and $c'_{24} = -0.16$ in unit of 10^{11} N/M^2, therefore c'_{24} can be neglected. In case of thickness shear mode, resonant frequency f is expressed by similar equations, and anisotropic constant becomes γ_{TS} given by

$$\gamma_{TS} = \sqrt{(c'_{22} - c'^2_{12}/c'_{11} + (\pi^2/12)c'_{66})/c'_{66}} \simeq 1.59,$$

where $c'_{22} = 2.64$, $c'_{11} = 2.28$, and $c'_{12} = 0.67$. Primed elastic constants c'_{66} etc are that of the rotated coordinate. Resonant frequencies are calculated from Eq. 5 and Eq. 6 as a function of a/H, and the results are shown in Fig. 2 as solid lines with the experimentally obtained data, where vertical shows $(f - f_e)/(f_o - f_e)$ and abscissa $(1/\gamma_{TT})(a/H)\sqrt{\Delta}$ or $(1/\gamma_{TS})(a/H)\sqrt{\Delta}$. And "⊙" sign corresponds to the case of thickness shear mode, "×" to thickness twist mode. In case of fundamental symmetric mode, theoretically obtained curve fits to the experimental data, howeverr, in case of antisymmetric mode and second symmetric mode the deviation become somewhat larger as the value of abscissa increasing.

In order to design a multimode monolithic filter, it is necessary to know the relation of bandwidth of pass band to dimension of electrode. We consider the simple case that there are only two modes, i.e., a fundamental symmetric and antisymmetric mode. In this case, band width is defined as

$$B = (f_{anti} - f_{symm})/\sqrt{f_{anti} \cdot f_{symm}} \quad , (9)$$

where f_{anti} and f_{symm} are the corresponding resonant frequencies. The relation B versus a/H are obtained from Table 3. and shown in Fig. 3. On the other hand, maximum band width is obtained from Fig. 2 as about 6 %, when a/H is equal to $0.35\gamma\sqrt{\Delta}$, i.e., in case of thickness shear mode a/H is equal to 1.85 and thickness twist mode to 1.04, respectively.

The coupling factor calculated from the physical constants is somewhat larger than the value obtained from resonant frequency, and resonant frequency lowering Δ depends on a propagating mode TS or TT. This is different from a case of quartz plate, where Δ is independent of a propagating mode in a given plate. From these fact, it is considered that the coupling factor depends on not only propagating mode but also the dimension of electrode. In case of specimen A the coupling factor is measured from the overtone frequencies of fundamental symmetric mode as a function of a/H. Fig. 4 shows that measured coupling factor approaches to some value below the k_t calculated from the physical constants with increasing a/H. From further experiments, coupling factor also depends on width (b) of electrode shown in Fig. 1. Therefore, it is necessary to consider propagating mode, length and width of electrode against thickness of plate in order to design a filter.

The mechanical quality factor Q is about 400 as shown in Table 2. This low Q were observed in all tested resonators vibrating thickness shear mode, and greatly depend on the holding condition. However, the Q values are greater than 10^4 in

case of the length expander bar[3]. The ultrasonic attenuation of $LiTaO_3$ measured by the author et.al.,[4] 0.07 dB/ μS at 50 MHz gives Q value probably greater than 10^4. The reason of this discrepancy is not clear, however, it may be thought that acoustic energy is mechanically dissipated in the trapping state.

Conclusion

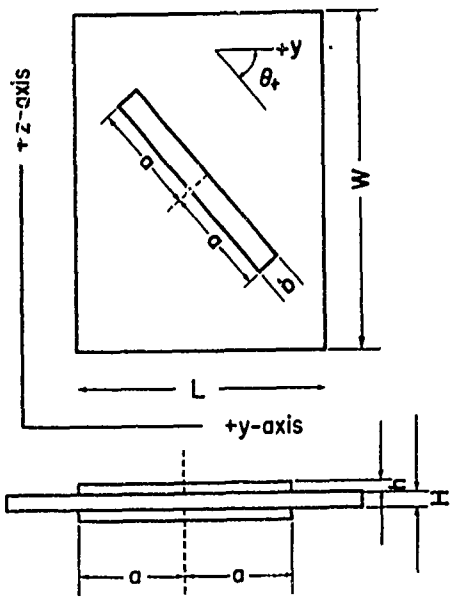
In a lithium tantalate X-cut resonator, resonant frequency lowering is mainly contributed by a piezoelectric effect and mass loading effect does not do its role. Considering a suitable anisotropic constant, resonant frequencies of thickness twist and thickness shear mode are expressed by a similar formula as a function of the ratio a/H, where a is the length of electrode and H the thickness of resonator. In case of multimode monolithic filter, obtainable maximum band width is about 6 % .

Acknowledgement

The author is indebted to Drs N. Niizeki and M. Hanzawa for their encouragement. He also wishes to thank Mr. T. Ashida for his helpful discussion.

References

1. M.Onoe, T.Ashida, and K.Sawamoto : Zero temperature coefficient of resonant frequency in an X-cut lithium tantalate at room temperature, Proc. IEEE vol. 57 pp 1446, 1969.
2. H. F. Tiersten : Linear Piezoelectric Plate Vibrations
3. K. Sawamoto, and N. Niizeki : Zero temperature coefficient of resonant frequency in $LiTaO_3$ length expander bars, Proc. IEEE vol. 58, pp 1289, 1970.
4. K. Sawamoto, T. Ashida, Y. Omachi, and T. Uno : Behavior of $LiTaO_3$ single crystal near its Curie point, J. Phy. Soc. Japan, vol. 28, Supplement, pp 309, 1970.
5. T.Uno: Direction of displacement of thickness shear mode transducer, J.Acoust. Soc. Japan, vol. 26, pp 399, 1970.



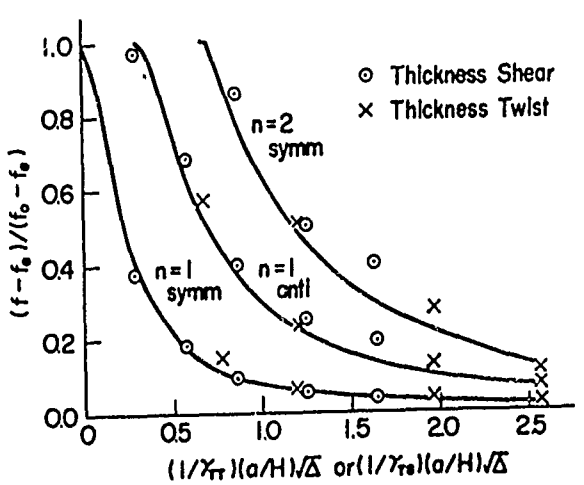
Dimmension of a lithium tantalate X-cut resonator, case of specimen A.

List of the characteristics of the specimens

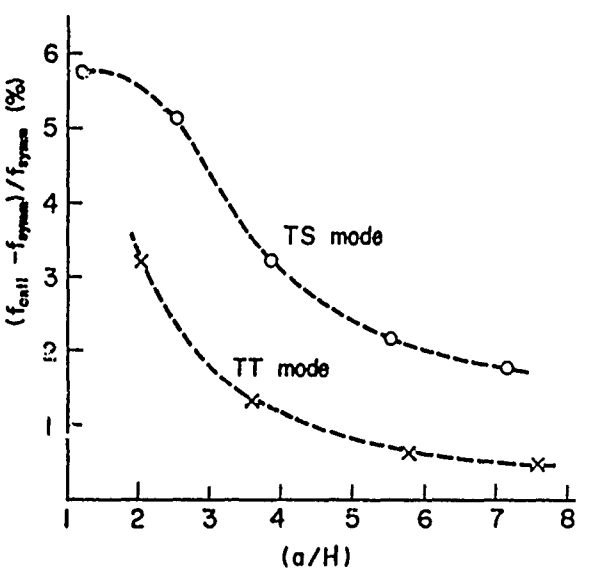
characteristic	specimen A	specimen B
L along y-axis	13.5 mm	13.7 mm
W along z-axis	17.5 mm	17.0 mm
H along x-axis	1.15 mm	0.94 mm
electrode length 2a	16.5 mm	14.4 mm
electrode width b	3 mm	3 mm
electrode thickness h	0.5 micron	0.15 micron
coupling factor	0.41	0.37
C_o	2.69 pF	
C_1	3.3 pF	
L_1	28 mH	
R	73.5 Ω	
Q	380	

Physical constants of $LiTaO_3$ X-cut plate

	fast shear wave	slow shear wave
velocity	$v_+ = 4.34 \times 10^5$ cm/s	$v_- = 3.39 \times 10^5$ cm/s
coupling factor	$k_f = 47.2\%$	$k_s = 5.5\%$
direction of displacement	$\theta_+ = -49°$	$\theta_- = +41°$



Relation of resonant frequency versus a/H.

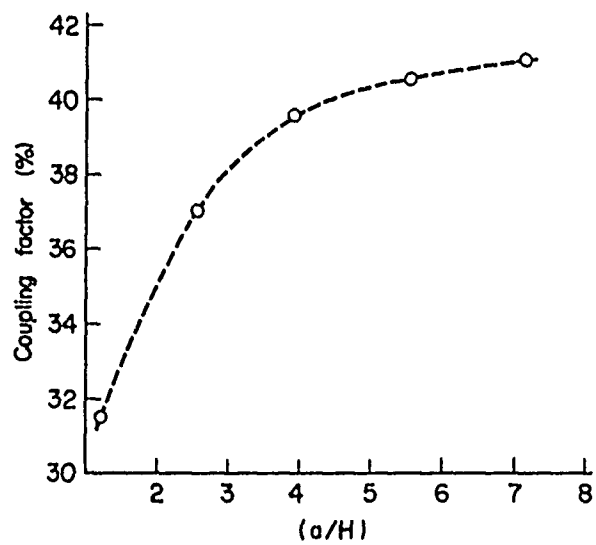


Band width versus (a/H)

Lengths of electrode a, and resonant frequencies.

specimen A	1	2	3	4	5
a mm	8.21	6.38	4.46	2.93	1.41
a/H	7.15	5.55	3.88	2.55	1.23
$f_{symm.}$ MHz fundamental	1.716	1.719	1.727	1.744	1.777
f_{symm} MHz second	1.784	1.803	1.866		
$f_{anti\text{-}symm}$ MHz fundamental	1.747	1.756	1.783	1.835	1.882
	f_0 = 1.890 MHz, f_0 = 1.710 MHz				

specimen B	1	2	3	4
a mm	7.18	5.49	3.93	1.94
a/H	7.68	5.84	3.61	2.06
$f_{symm.}$ MHz fundamental	2.144	2.148	2.151	2.170
$f_{symm.}$ MHz second	2.171	2.189	2228	
$f_{anti\text{-}symm}$ MHz fundamental	2.155	2.162	2.180	2.241
	f_0 = 2.310 MHz, f_0 = 2.140 MHz			



Coupling factor versus (a/H)

Resonant frequency lowering, mass loading, and piezo-electric effect

	specimen A	specimen B
coupling factor k	0.41	0.37
frequency lowering Δ	9.0 %	7.4 %
mass loadig effect R	0.051 %	0.019 %
piezoelectric effect P	7.4 %	6.1 %

CdS-QUARTZ MONOLITHIC FILTERS FOR THE 100-500 MHz FREQUENCY RANGE

D. A. Roberts
Gould Inc.
Gould Laboratories - Cleveland
Cleveland, Ohio

Summary

A severe limitation on quartz monolithic filter bandwidth is imposed by a reduction in the effective electromechanical coupling factor at harmonic modes which are necessary for filter applications above 100 MHz. This paper describes a method by which the upper bandwidth limitation can be extended nearly two orders of magnitude by utilizing a high coupling piezoelectric thin film (CdS) in combination with the conventional (~ AT) quartz filter crystal. The composite CdS-quartz filter can be designed to have nearly the frequency-temperature stability of AT-quartz alone (< 1 ppm/°C) and can be terminated directly in a small resistive impedance without the need for additional components such as coils or transformers. Energy trapping principles and filter design methods apply, just as with conventional quartz filters.

Measurements on a number of CdS-quartz resonators and filters are reported with emphasis on results for 190 MHz 7th harmonic mode 2, 4 and 6-pole filter prototypes. Low ripple Chebyshev and maximally flat bandpass filter responses are achieved with these units, having 3 dB bandwidths around 130 kHz, 1-3 dB insertion loss per resonator and > 50 dB ultimate out-of-band rejection per 2-resonator filter section when terminated directly in 50 Ω. Direct coupled 2-resonator filter sections yield 4 and 6 resonator tandem monolithic filter configurations which are shown to be effective in suppressing spurious modes as well as increasing ultimate out-of-band attenuation and filter selectivity.

Design formulae are presented for fabricating CdS-quartz composite resonators to predetermined specifications (k_{eff} and Q_M) for subsequent use in filter design. Experimental and theoretical comparisons between quartz and CdS-quartz devices are discussed.

Introduction

The conventional quartz filter crystal is currently limited to fundamental mode frequencies around 40-50 MHz. Higher fundamental frequencies have been realized but require quartz wafers so thin as to border on the limits of present day manufacturing capabilities. It is possible, of course, to operate a quartz resonator at odd harmonics (3, 5, 7, ...) above the fundamental, but under [illegible]ed conditions the effective electromechanical coupling factor k_{eff} is reduced to $1/p$ times its value for fundamental mode operation. Since the maximum attainable inductorless bandwidth for a crystal filter is proportional to k^2_{eff}, wave filters utilizing conventional overtone quartz resonators are severely limited in bandwidth above 100 MHz. [1]

The device to be described in this paper is a thin film cadmium sulfide-quartz crystal filter which combines the high frequency-temperature stability of AT-quartz with significantly higher electromechanical coupling than is possible with a comparable quartz crystal filter for the same frequency and wafer thickness. It is anticipated that the composite filter will have its greatest utility for frequencies in the range 100-500 MHz.

Before discussing the composite CdS-quartz filter, it is instructive to review the essential features of the thin film composite resonator, shown schematically in Fig. 1. [2] The device consists of an oriented thin film piezoelectric transducer (2-5 μ) which is deposited with metal electrodes onto a high Q wafer substrate (typically 20-30 MHz fundamental frequency). This structure can be operated at any number of even or odd harmonic modes, but the strongest responses (i.e., responses having lowest motional impedance) result for those harmonics at which there is roughly one-quarter to one-half acoustic wavelength in the transducer film. Energy trapping techniques can be applied in order to confine the vibratory energy to the CdS-quartz volume in the immediate vicinity of the electroded region, just as with conventional quartz resonators. [3] The electrical equivalent circuit for a composite resonator near a strong resonance response is likewise identical to that used to represent its quartz counterpart (Fig. 1). The shear mode version of the composite resonator can be designed to have a high degree of frequency stability over any specified range of temperatures by utilizing a Y-rotated cut of quartz close to the AT or BT cuts having a small positive temperature coefficient which balances the negative coefficient of the CdS layer. [4]

In Fig. 2, electromechanical coupling k_{eff} and mechanical Q are compared for (1) a conventional quartz resonator, (2) a CdS-quartz resonator having a quarter-wave transducer and (3) a CdS-quartz resonator having a half-wave transducer. Numerical examples are also given comparing shear mode designs near 300 MHz assuming that the same AT-quartz wafer (33 MHz fundamental frequency) is used in each design. The formulae and examples given illustrate the principle dynamic feature of the composite r[illegible] which is high electromechanical coupling at overtone frequencies. Corresponding maximum inductorless fractional bandwidths ($\approx k^2_{eff}/2$) are given in the right hand column of Fig. 2 for bandpass filters utilizing each of the three resonator design examples, thus demonstrating the theoretical capability for achieving up to 100 times larger bandwidths with CdS-quartz monolithic filters than is possible with quartz alone. It is also important to note that, because of the moderately large values of k_{eff} for a composite resonator and because the transducer electrodes are separated only by the very thin transducer film, the motional impedance of a composite resonator is considerably smaller than for a comparable quartz resonator geometry. Two practical advantages are gained by this feature. (1) much smaller electrodes may typically be used in composite resonator (or filter) design without being forced to intolerably high impedance levels, thus easing the problem of complying with energy trapping theory for strong, spurious free responses while maintaining reasonable electrode thickness and (2) small filter termination resistances (50-300 Ω) are possible for typical

composite filter designs, which virtually eliminates the need for balancing out inevitable parasitic capacitances with external coils.

Design and Fabrication

The geometry of a 2-resonator CdS-quartz monolithic filter is illustrated in Fig. 3 together with typical dimensional values for filters which have been successfully fabricated. The design consists of two acoustically coupled composite resonators on the same quartz wafer. The design process is carried out by choosing a composite resonator configuration (using Fig. 2, for example) for which k_{eff} and Q_M satisfy bandwidth and insertion loss specifications for the filter insertion loss characteristic desired (Butterworth, etc). Then electrode and electrode gap dimensions and plate-back are calculated according to ordinary overtone quartz design methods, consistent with the desired filter termination impedance and normalized filter design parameters tabulated in filter design handbooks. In other words, once the composite resonator structure is decided, all else proceeds just as for quartz overtone filter design.[5,6,7] A possible exception is that the quartz coupled resonator theory might need slight alteration to account for the presence of the CdS transducer layer. For the 190 MHz 2-pole filter design reported in this paper, no such alterations were necessary.

The fabrication of these filters consists first of depositing split aluminum electrodes onto the quartz wafer substrate. For the design reported here the electrodes are oriented so that the inter-resonator coupling direction is the quartz Z'-axis, thus utilizing quartz thickness twist waves. Next, the wafer is heated to about 180°C and a CdS layer is vacuum deposited from an oblique angle of 50° from the wafer normal at a rate of 0.2 μ/min. Separate Cd[1] and sulfur evaporation sources are used in an evaporator system similar to one described by N. F. Foster.[8] The orientation of the deposited polycrystalline CdS film is such that the hexagonal c-axes make an average angle of about 30° with the substrate normal (maximum piezoelectric coupling to the quasi-shear mode), and the projection of the CdS c-axes onto the plane of the wafer coincides with the quartz X-axis.[2] Before cooling the CdS-quartz structure back to room temperature, a thin layer of copper or copper sulfide is deposited over the CdS in order to help compensate excess sulfur vacancies in the film. Finally, the split top aluminum electrodes are deposited onto the room temperature composite structure. The evaporator vacuum is released between successive metal and CdS depositions in order to change evaporation masks. A photograph of a mounted CdS-quartz filter is shown in Fig. 4 together with a simple test fixture for making filter insertion loss measurements. The test fixture is designed to present 50 Ω terminations at the filter package terminals when connected into 50 Ω measurement circuitry. For 2-terminal resonator measurements, the printed circuit connections on the backside of the wafer mounting board permit the two (coupled) resonators to be connected electrically as a single resonator which can then be excited either in the symmetric mode (resonators vibrating in phase) or the antisymmetric mode (resonators vibrating 180° out of phase).

Experimental Results

Before proceeding to the filter development results, it is instructive to observe the overall frequency response of a CdS-quartz composite resonator. A half-lattice (hybrid coil) measurement bridge is used to plot the response shown in Fig. 5 for resonator CR03-D.[9] The curve demonstrates the application of conventional energy trapping principles to achieve strong, spurious-free responses in CdS-quartz resonators. For quantitative resonator measurements, complex impedance vs. frequency is measured using a conventional VHF impedance bridge with special fixturing for tuning out stray reactances.[4] The data is then plotted on polar impedance coordinates (Z, θ), as shown in Fig. 6. In this example, the impedance characteristic is measured for the (coupled) resonators of a monolithic filter in which the two resonators are connected together as a single resonator in the antisymmetric mode excitation. The impedance circle plot makes possible the determination of resonator parameters such as motional resistance at resonance R_1, shunt capacitance C_o, electromechanical coupling k_{eff} and mechanical quality factor Q_M.[4,10] A summary of measured resonator parameters for a number of CdS-quartz resonators and filters is given in Fig. 7. In the right hand column, harmonic mode p is given along with an indication whether the response measured is a symmetric or antisymmetric mode. The first two entries, PR-14C and CR03-B, are single resonator units with circular electrodes, so they are of necessity excited in the symmetric mode. These two resonators and the first filter entry, CF01-B, have half-wave ($\lambda/2$) CdS transducers at the harmonic modes indicated, while the remaining filter entries have quarter-wave ($\lambda/4$) transducers The socalled resonator "figure of merit" $k_{eff}^2 Q_M$ is in the neighborhood of 8 for all these units; hence, one can predict at once that 2-pole CdS-quartz bandpass filters with the resonator parameters given in Fig. 7 and having Butterworth or low ripple Chebyshev insertion loss characteristics should have insertion loss values in the neighborhood of 4 dB (not including losses due to series electrode resistance) for properly terminated and optimally designed units.[5] It is interesting that state-of-the-art conventional quartz resonators operating at comparable harmonics and frequencies exhibit resonator "figures of merit" comparable to or slightly lower than those reported in Fig. 7, although the individual k_{eff} and Q_M values differ considerably from corresponding values for CdS-quartz resonators.[3,11]

The remainder of this paper describes the successful utilization of CdS-quartz resonators in composite filter fabrication. Figures 8 and 9 give attenuation and phase delay vs. frequency for a 191 MHz, 7th harmonic mode monolithic filter which is terminated directly in 50 Ω transmission line. The passband insertion loss can be reduced to 4 dB while at the same time virtually eliminating passband ripple by terminating the filter in a load slightly larger than 50 Ω. The 135 kHz 3 dB bandwidth is close to the optimum design fractional bandwidth given by $k_{eff}^2/2 = 0.08\%$ and compares to the maximum inductorless bandwidth for quartz at this harmonic, $k_{eff}^2/2 \simeq 0.007\%$. An outstanding feature of the insertion loss characteristic in Fig. 8 is the > 50 dB ultimate out-of-band rejection, impressive for a 2-pole filter at this frequency. The small spike near 182 MHz results from slightly inaccurate alignment of the hexagonal c-axes of the polycrystalline CdS film with respect to the quartz X-axis, so that a second shear mode series (the socalled "fast" shear modes, controlled primarily by the quartz elastic constant c_{44}) is electrically excited in the composite structure in addition to the main harmonic mode series (socalled "slow" shear series). In fact, if the

frequency scale in Fig. 8 were extended far enough, one could observe not only two thickness shear mode harmonic series, but also a thickness extensional mode series which is piezoelectrically excited by virtue of our choice of CdS c-axis orientation with respect to the quartz wafer normal.[2, 13] The small spike at 182 MHz is, of course, well suppressed relative to the main (slow shear) harmonic mode of interest, but it is possible that the stronger thickness extensional modes or even the other very strong slow shear harmonics (also a problem with overtone quartz filters) would yield undesirable additional passbands. Unwanted harmonic modes as well as spurious (or inharmonic) modes associated with the main mode of interest can be greatly suppressed while at the same time increasing ultimate out-of-band rejection and filter selectivity by connecting two or more 2-pole filter sections in tandem, each filter section being designed for a different harmonic (i.e., p = 7 and p = 8) such that only the main modes of interest coincide in frequency. This technique is particularly effective using composite filters because ultimate out-of-band rejection for individual 2-pole filters is so high and because, unlike quartz filters, adjacent harmonics having between one-quarter wave and one-half wave in the transducer layer have almost indistinguishable motional parameters.[2]

Figure 10 shows the insertion loss characteristic for a 4-pole tandem monolithic filter in which two identical 2-pole filter sections are direct-coupled together (no interstage reactive elements), thus preserving the truly monolithic nature of the CdS-quartz filter. This filter is terminated in slightly greater than 50 Ω impedance in order to achieve a maximally flat (Butterworth) passband characteristic. The 30 dB/3 dB skirt ratio for this unit is 2.4, very close to the value expected theoretically.[5] In Fig. 11 a third 2-pole filter section has been inserted in tandem with the 4-pole filter just discussed, giving a 6-pole filter characteristic with excellent selectivity. The 6-pole filter is not a legitimate minimum insertion loss design because three 2-pole sections having identical center frequencies and bandwidths are simply direct-coupled together, whereas optimum insertion loss design specifies non-identical sections for a 6-pole response.[5] This consideration explains the unusually high insertion loss for the response shown in Fig. 11.

Figures 12 and 13 demonstrate the utilization of a 4-pole CdS-quartz tandem monolithic filter in a circuit application, in this case a filter-amplifier circuit. Figure 12 shows that the circuit design utilizes only transistors, capacitors and resistors which are compatible to hybrid chip (IC) realization in order to be consistent with the truly monolithic nature of the composite filter. Although some gain is sacrificed, the need for coils has been completely eliminated, a feature which could not be achieved using conventional broad-banded quartz crystal filters at this frequency and bandwidth. The 10 dB pass band gain shown in Fig. 13 compares to 12 dB passband insertion loss for this particular 4-pole filter when terminated in 50 Ω. Bandwidth and center frequency are virtually unaffected by the presence of the surrounding circuitry, although some deterioration in ultimate out-of-band rejection did result from inadequate shielding and grounding in the chassis construction.

Temperature characteristics for CdS-quartz filters operating at half-wave and quarter-wave transducer modes are shown in Fig. 14 (indicated as Types I and II, respectively). Both filters were designed to have optimum frequency stability over the -40°C to +85°C temperature range when operated at the half-wave transducer mode.[5] The data presented in Fig. 14 demonstrate the high degree of stability possible with the composite CdS-quartz filter. The maximum filter center frequency excursion from room temperature frequency is only a few percent of filter 3 dB bandwidth for these filter designs.

The extension of the present work to lower frequencies than 190 MHz is straightforward and should yield low loss multi-pole filter designs. The extension to frequencies higher than 190 MHz has been attempted on a limited basis. CdS-quartz filters have been operated as high as 400 MHz, giving flat passbands and good selectivity. Figure 15 shows a 335 MHz 2-pole filter response utilizing the 11th thickness extensional harmonic mode, having a 3 dB bandwidth of 390 kHz and minimum insertion loss of 10 dB.[12] Shear mode filter development above 300 MHz has so far yielded units with even higher insertion loss. It is felt that further filter design optimization together with modest improvement in resonator properties (i.e., $k^2_{eff}\, Q_M$) will be necessary before low loss composite filters can be fabricated at frequencies above 300 MHz. This may require a more suitable transducer material than thin film CdS, but it should be pointed out that the CdS film properties achieved in this work, i.e., $k_{TS}(CdS) \leq 0.17$ and $Q_M(CdS) \leq 700$ at 200 MHz, are not necessarily indicative of the parameters values which could be realized in an optimized film deposition system.

Acknowledgments

The work reported in this paper was supported by the United States Army Electronics Command, Fort Monmouth, New Jersey, under Contract DAAB07-69-C-0356. The author would like to acknowledge helpful assistance from his colleagues G. Conley and D. J. Koneval and also former colleagues K. A. Pim and D. A. Berlincourt, now with the Piezoelectric Division, Vernitron Corp., Bedford, Ohio.

References

1. R. C. Smythe, "HF and VHF Inductorless Filters for Microelectronic Systems," Proc. 1969 Electronic Components Conf., Washington, D.C., pp. 115-119.

2. T. R. Sliker and D. A. Roberts, "A Thin Film CdS-Quartz Composite Resonator," J. Appl. Phys. 38, 2370 (1967).

3. W. Shockley, D. R. Curran, and D. J. Koneval "Trapped-Energy Modes in Quartz Filter Crystals," J. Acoust. Soc. Am. 41, 981(1967).

4. D. A. Roberts, "UHF Filter Research," Final Report, Contract DAAB07-69-C-0356, U.S. Army Electronics Command, Ft. Monmouth, N.J. (May or June, 1971).

5. A. I. Zverev, Handbook of Filter Synthesis, New York, Wiley, 1967

6. M. Onoe and H. Sumonji, "Analysis of Piezoelectric Resonators Vibrating in Trapped Energy Modes," Electronics and Comm. Eng. (Japan), Vol. 48, No. 9, 84-93, September, 1965.

7. R. A. Sykes, W. L. Smith and W. J. Spencer, "Monolithic Crystal Filters," Presented at IEEE Int'l Conv., New York, March 22, 1967.

8. N. F. Foster, G. A. Coquin, G. A. Rozgonyi and F. A. Vannatta, "Cadmium Sulphide and Zinc Oxide Thin-Film Transducers," IEEE Trans. on Sonics and Ultrasonics SU-15, 28 (1968).

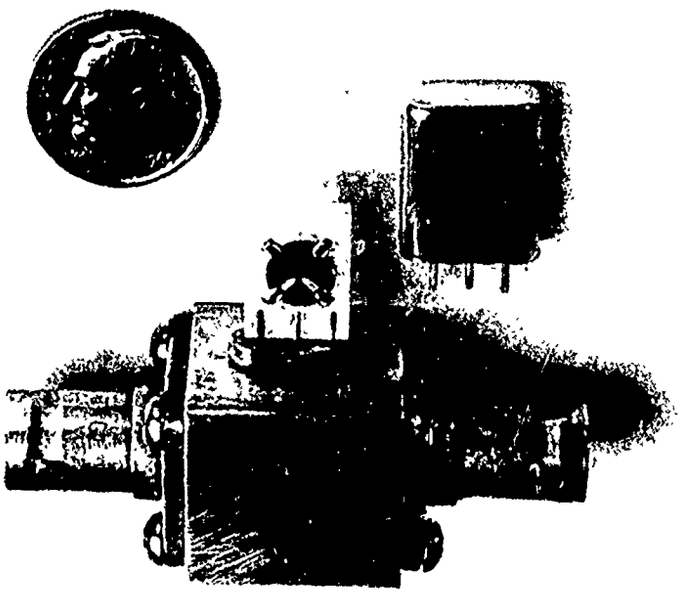
9. W. H. Horton and R. C. Smythe, "The Hybrid-Coil Bridge Method of Measuring Unwanted Modes of Vibration in Quartz Crystals," Proc. 17th Ann. Freq. Control Symp., May 1963.

10. Richard Holland and Errol P. Eernisse, "Accurate Measurement of Coefficients in a Ferroelectric Ceramic," IEEE Trans. on Sonics and Ultrasonics SU-16, No. 4, 1969, pp. 173-181.

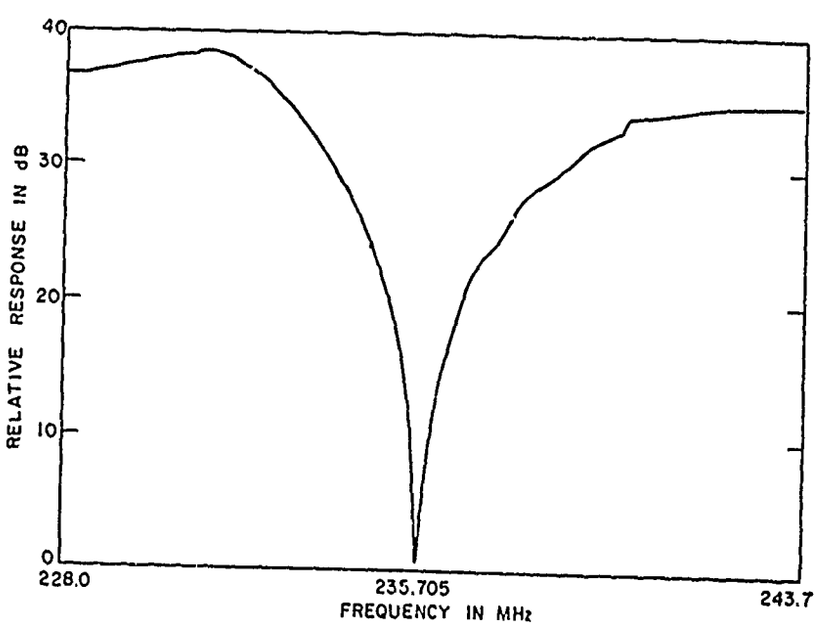
11. D. J. Koneval and T. R. Sliker, "VHF Filter Crystals Research," Final Report, Contract DAAB07-67-C-0240, U.S. Army Electronics Command, Ft. Monmouth, N.J. (Jan., 1969).

12. D. A. Roberts, "Hybrid Piezoelectric Devices," Final Report, Contract DAAB07-67-C-0163, U.S. Army Electronics Command, Ft. Monmouth, N.J. (November, 1969).

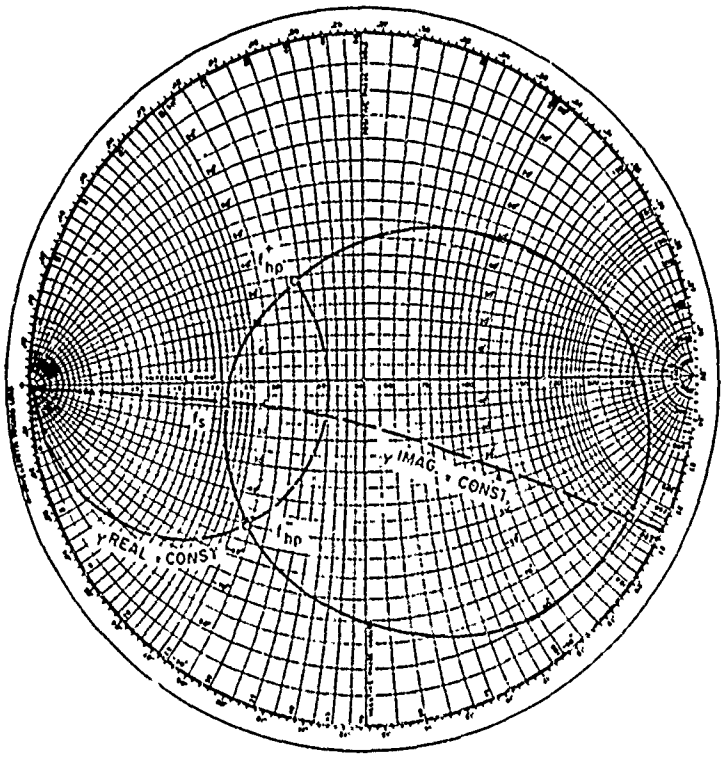
13. Walter G. Cady, Piezoelectricity, (Dover Publications, Inc., New York, 1964).



Mounted CdS-Quartz Monolithic Filter and Test Fixture



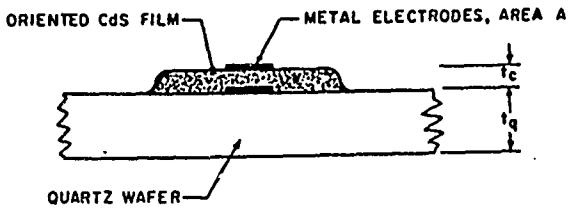
HALF-LATTICE BRIDGE RESPONSE FOR CdS-QUARTZ RESONATOR CRO3-D AT THE 9th SLOW SHEAR HARMONIC MODE

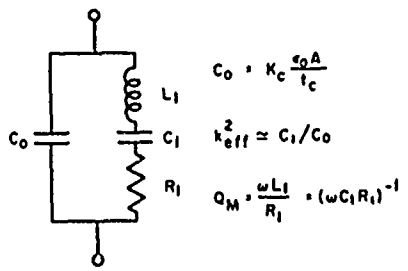


RESONATOR Z-θ IMPEDANCE CIRCLE FOR THE FIRST ANTISYMMETRIC TRAPPED ENERGY MODE IN CdS-QUARTZ COMPOSITE MONOLITHIC FILTER CFO5-D AT THE 7th THICKNESS SHEAR HARMONIC

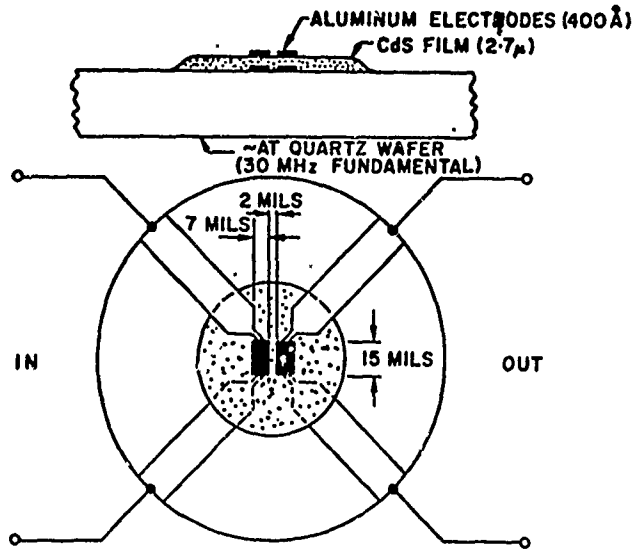
SUMMARY OF MEASURED RESONATORS PARAMETERS FOR SELECTED CdS-QUARTZ RESONATORS AND FILTERS

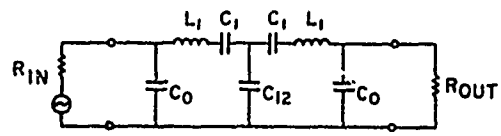
UNIT	f(MHz)	C_0(pF)	k_{eff}	Q_M	$k^2_{eff}Q_M$	MODE
PR-14C	234	3.5	5.2 %	3400	9.2	p=9, SYMM
CR03-B	235	3.9	5.2 %	4000	10.8	p=9, SYMM
CF01-B	191	1.5	6.1 %	2100	7.8	p=7, SYMM
CF05-B	191	2.2	3.6 %	5700	7.4	p=7, ASYM
CF05-D	191	2.1	4.0 %	6300	10.0	p=7, ASYM
CF06-B	191	2.3	3.4 %	6200	7.2	p=7, ASYM
CF07-D	191	2.1	3.2 %	7000	7.0	p=7, ASYM
CF09-B	191	2.2	2.6 %	9300	6.3	p=7, ASYM





SCHEMATIC OF CdS-QUARTZ COMPOSITE RESONATOR WITH EQUIVALENT ELECTRICAL CIRCUIT VALID AT ANY OF ITS STRONG (EVEN OR ODD) HARMONIC MODES.



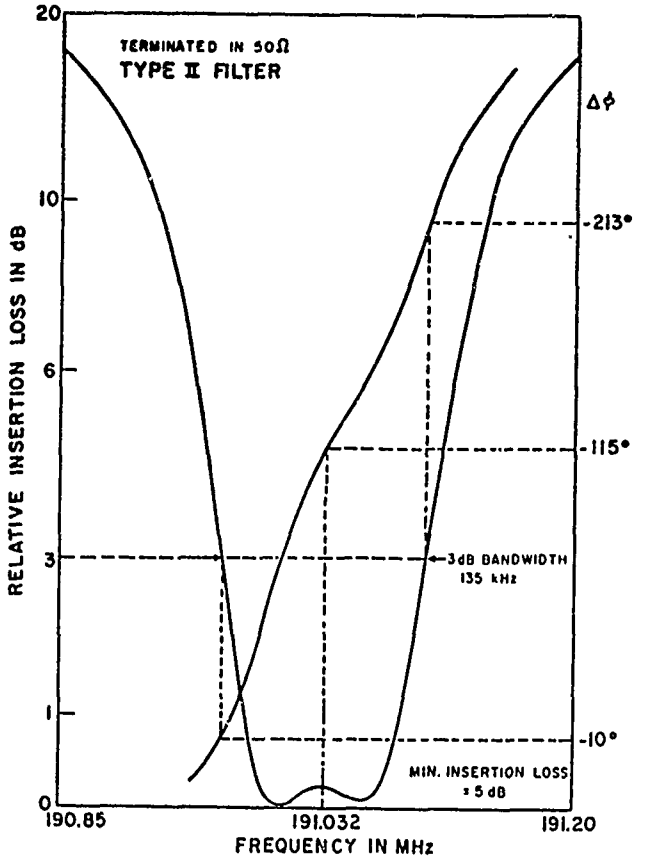


SCHEMATIC OF 2-RESONATOR CdS-QUARTZ MONOLITHIC FILTER WITH EQUIVALENT CIRCUIT FOR A 2-POLE BANDPASS FILTER.

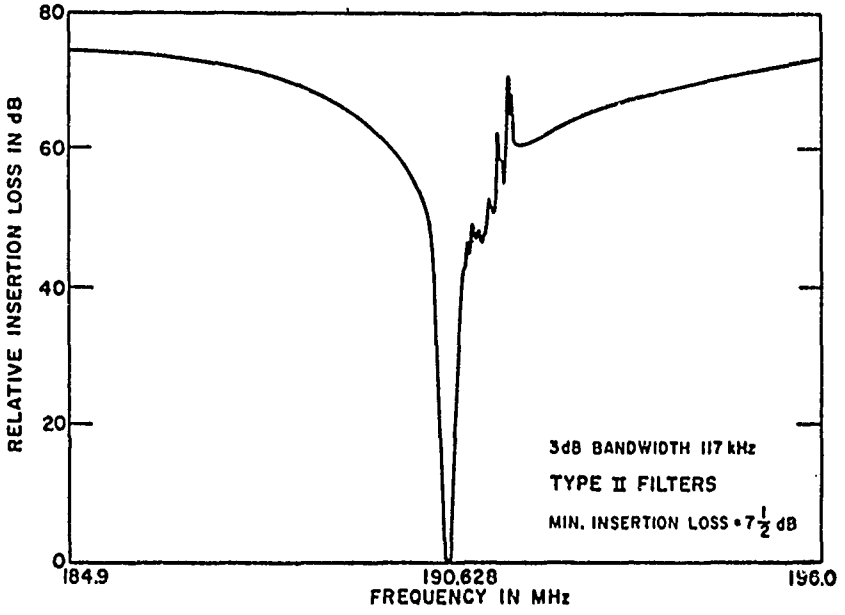
COMPARISON OF QUARTZ AND CdS-QUARTZ RESONATORS*

	k_{eff}	Q_M	MAXIMUM INDUCTORLESS FRACTIONAL BANDWIDTH
QUARTZ	$\sqrt{\frac{8}{\pi^2}}\ \frac{k_q}{p}$ (0.9 %)	Q_q (30,000)	(0.004 %)
CdS-QUARTZ $\lambda/4$ CdS $m\lambda/4$ QUARTZ	$\sqrt{\frac{8}{\pi^2}}\ \frac{k_c}{\left[1 + m Z_c/Z_q\right]^{1/2}}$ (4.5 %)	$\left[\frac{1}{Q_q} + \frac{1}{m}\frac{Z_q}{Z_c}\frac{1}{Q_c}\right]^{-1}\left[1+\frac{1}{m}\frac{Z_q}{Z_c}\right]$ (8300)	(0.100 %)
CdS-QUARTZ $\lambda/2$ CdS $n\lambda/2$ QUARTZ	$\sqrt{\frac{8}{\pi^2}}\ \frac{k_c}{\left[1 + n Z_q/Z_c\right]^{1/2}}$ (7.0 %)	$\left[\frac{1}{Q_q} + \frac{1}{n}\frac{Z_c}{Z_q}\frac{1}{Q_c}\right]^{-1}\left[1+\frac{1}{n}\frac{Z_c}{Z_q}\right]$ (4000)	(0.250 %)

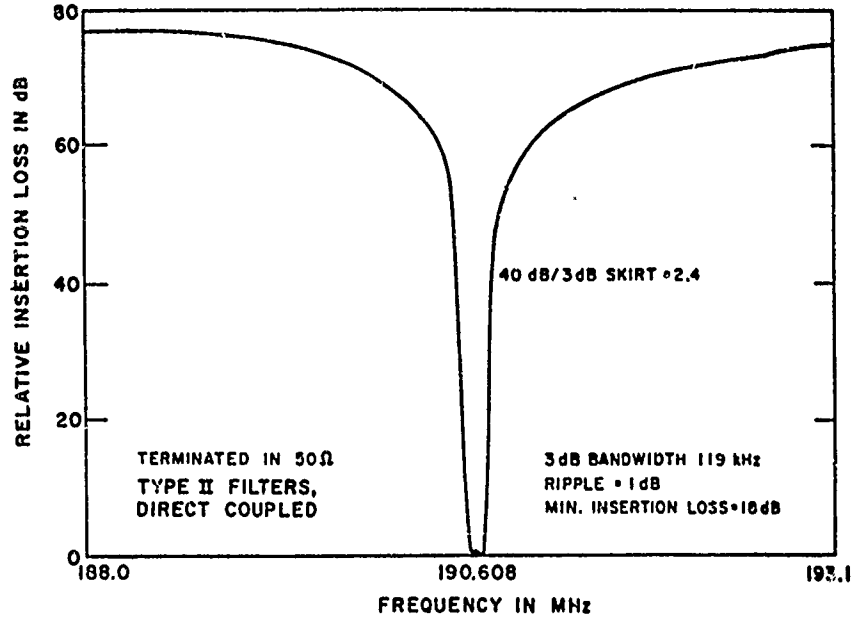
*NUMERICAL EXAMPLES IN PARENTHESES ASSUME 33 MHz FUNDAMENTAL AT-QUARTZ WAFERS ARE USED AND THAT EACH RESONATOR IS OPERATED AT THE HARMONIC NEAREST 300 MHz.



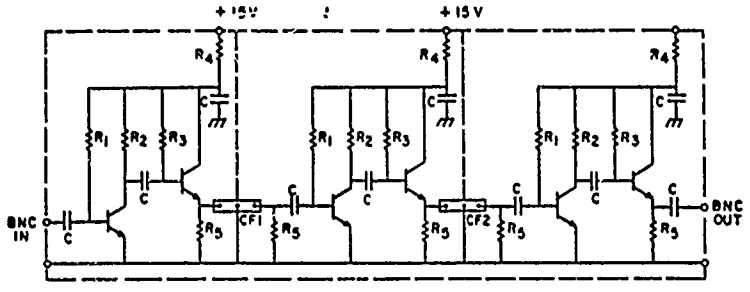
CdS-QUARTZ COMPOSITE MONOLITHIC FILTER CF05-D AT THE 7th THICKNESS SHEAR HARMONIC MODE SHOWING PASSBAND ATTENUATION AND PHASE CHARACTERISTICS.



CdS-QUARTZ TANDEM COMPOSITE MONOLITHIC (4-POLE) FILTER CF06-B/CF05-D AT THE 7th THICKNESS SHEAR HARMONIC MODE. TERMINATION RESISTANCES ADJUSTED FOR FLAT PASSBAND.

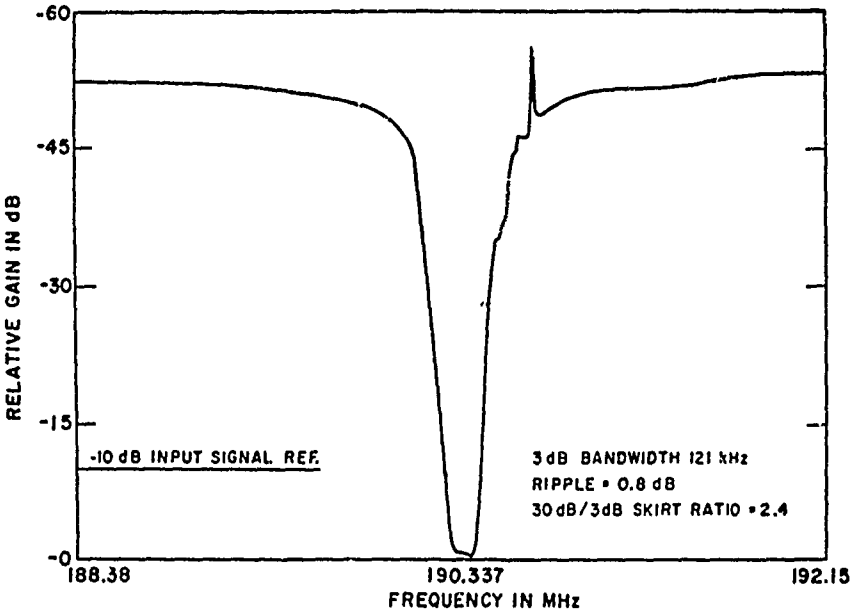


CdS-QUARTZ TANDEM COMPOSITE MONOLITHIC 6-POLE FILTER CF06-B/CF07-D/CF05-D AT THE 7th SLOW SHEAR HARMONIC MODE WHEN TERMINATED IN 50Ω LINE.

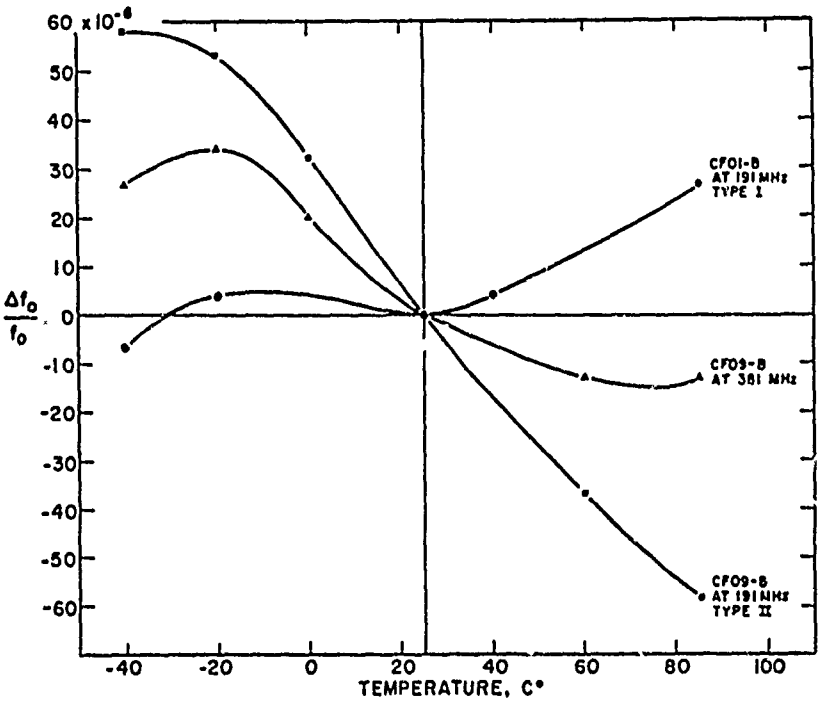


R_1 68K, 10%, 1/4 W	ALL CAPACITORS: .01μfd, 20V, CERAMIC
R_2 470, 10%, 1/4 W	ALL TRANSISTORS: 2N3866 (SHOULD BE HEAT SINKED)
R_3 100K, 10%, 1/4 W	CF1, CF2 = 2-POLE CdS-QUARTZ MONOLITHIC FILTERS, 50Ω DESIGN.
R_4 10Ω, 10%, 1/4 W	MINIMUM GAIN: 6 dB/STAGE
R_5 50Ω, 1%, 1/4 W (METAL FILM)	POWER INPUT: 85 mA AT +15V

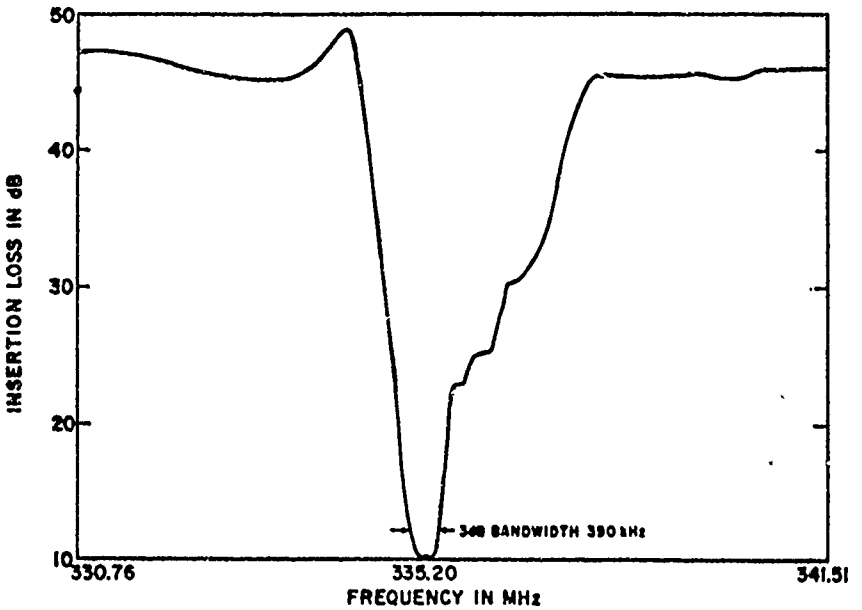
190 MHz FILTER-AMPLIFIER UTILIZING CdS-QUARTZ MONOLITHIC FILTERS.



VHF FILTER-AMPLIFIER GAIN vs. FREQUENCY CHARACTERISTIC. CdS-QUARTZ COMPOSITE MONOLITHIC (2-POLE) FILTERS CF06-D & CF05-B WERE USED TOGETHER WITH THREE STAGES OF AMPLIFICATION--NO INDUCTORS OR TRANSFORMERS.



VARIATION IN BANDPASS CENTER FREQUENCY f_0 WITH TEMPERATURE FOR TYPE I CdS-QUARTZ FILTER CF01-B AND TYPE II CdS-QUARTZ FILTER CF09-B



FILTER RESPONSE FOR CdS-QUARTZ COMPOSITE COUPLED-MODE. FILTER 08IA-M AT THE 11th THICKNESS EXTENSIONAL HARMONIC MODE.

CONSIDERATIONS ABOUT CHANNEL FILTERS FOR A NEW CARRIER FREQUENCY SYSTEM WITH MECHANICAL FILTERS

by

H. Schüssler

AEG-TELEFUNKEN Forschungsinstitut, 79 Ulm/Do. (W-Germany)

Abstract

A premodulation system with integral signalling which needs only one type of channel and signalling filters each, has been developed. Mechanical filters with torsional vibrating resonators at a carrier frequency of 200 kHz fulfill the desired specifications (1/10 CCITT). The channel filter consists of 6 and the signalling filter of 3 resonators. Additional bridging wires allowed the reduction of the number of resonators by finite poles. Therefore the volume of the channel filter could be reduced to 6 cm^3.

Fully automatic production of the torsional filters will be possible. The main production steps, material problems and the influence of both of them on the transmission properties of the filters are discussed. With the proposed filters it is possible to accommodate 600 - 900 channels in a bay.

Modulation System

The growing market for telephone equipment compels manufacturers to design more automatic and economic systems and production techniques.

At the present time two carrier frequency systems are used: one of them without integral signalling in a kind of direct modulation with 12 types of crystal filters, the other with integral signalling in a pregroup system with 8 types of coil filters.

The earlier systems had been chosen to minimize the number of needed carriers and modulators. The reliability, the small volume and lower power consumption of semiconductor equipment allows increasing the number of channels for one carrier supply. The expense for modulators has also become smaller. Therefore, it seems interesting to minimize the expense for filters.

Crystal filters [1,2] or coil filters[3] could not be replaced economically by mechanical filters without changing the system. Premodulation systems with only one type of filter or 12 types which differ only slightly, have been proposed[4,5,6].

Figure 1 shows the frequency plan and a circuit diagram of the channel unit for a premodulation system. The VF band is first modulated with a 200 kHz carrier and filtered by the channel filter. The signalling carrier with 203,825 kHz is then added. In a second step the channels with integral signalling are modulated with 12 different carriers from 264 to 308 kHz to form a primary group. The upper sideband and other intermodulation products are supressed by one type of passive RC-filter. The channel unit consists of 2 channel filters, one signalling filter at the receiver end, 3 lowpass filters, 5 modulators and a signal receiver. The lowpass filter LP1 should only form the signal, and the other filters LP2 and LP3 suppress signals of broadcasting transmitters and jamming modulation products. All these filters are simple RC networks.

The choice of carrier frequency for the premodulation depends on the filter concept. System engineering shows that below the primary group there exists only the premodulation frequency of 48 kHz without jamming modulation products, and the carriers for the second modulation spaced far enough from the band of the primary group. Above 130 kHz there are no difficulties from the system, but the specifications for the undesired couplings and the stability of components are more stringent.

Filter Approximation

The tolerances for transmission systems are fixed by CCITT. The tolerance for a single filter is only a small part of the overall tolerance. In Fig. 2 several transmission curves are shown for two types of channel and one type of signalling filter. Due to the integral signalling for the channel filter the same shape factor is needed on the carrier end upper side of the transmission curve. A symmetric approximation can be used.

A polynomial filter with 12 zeros and a filter with 3 pairs of finite poles and 8 zeros, both having a reflexion coefficient of 10 %, fulfill the specification if the factor Q is high enough. The transmission curves in the passband show that the normalized q, defined by the following equation

$$q = Q \cdot B_3 / f_M \qquad (1)$$

with the 3 dB bandwidth B_3 and the centre frequency f_M of the filter, must be higher than 250 for a channel filter. For smaller distortions in the passband (1/20 CCITT) the q must be higher than 400.

The polynomial filter needs a higher q than the filter with damping poles due to the stronger edge distortion. An approximation with predistortion, however, shows greater sensitivities with respect to production tolerances and will not be discussed further. Monolithic crystal filters with a carrier frequency of 8140 kHz and a q = 80 can not satisfy the specification for a system with integral signalling. The proposed system with monolithic crystal filters works without integral signalling and therefore allows greater attenuation distortion in the passband and demands lower attenuation values in the stopband[7].

The signalling filter for 203,825 kHz is calculated from a Butterworth approximation with 100 Hz bandwidth. With q = 250 the transmission losses are in the order of 3 dB. Channel filter and signalling filter transmission curves are drawn in the stopband.

The stability is the other important argument for a filter concept. It is assumed that the centre frequency of the channel and signalling filters may change less than 10 Hz in the temperature range 10^{0}C to 50^{0}C and less than 10 Hz in the whole life of the filters. In Fig. 3 the comparison of the Q , the temperature coefficient and the aging rate per year under these conditions for four filter concepts are given. The values in brackets are experimental values which we have found in publications or which we have measured ourselves.

On account of the too low q values a premodulation concept with LC-filters concentrated in a block and monolithic crystal filters cannot be realized. Temperature coefficient and aging rate are at an extreme point in this case also. The two concepts with mechanical filters have enough reserves.

Mechanical Filters

A lot of proposals have been made for filter structures with mechanical resonators[8]. An optimal concept must be especially measured on the production expenses. Earlier concepts are not satisfying. There are two new concepts for mechanical channel filters which seem to be interesting.

For a carrier frequency of 48 kHz a polynomial filter with bending resonators and transducers and only one longitudinal vibrating coupling wire has been described[9]. Figure 4 shows a drawing of such a filter that has been obtained from a patent application. The bending resonators (6) are approximetely 17 mm long and have a cylindrical form with a flat face on one side so that there is a fixed orientation of the mode. For supporting the resonators two bars (4) are fixed at one end in the nodes of the bending resonators and at the other end soldered in the bottom of the case. The transversally excited transducers consist of a metal part (6') and a disc of PZT ceramic (10). The polarisation of the bending vibration is in the filter plane. The length of the longitudinally vibrating coupling wire (8) between two resonators is in the order of $\lambda/20$ and the wire has a thickness of approx. 0,5 mm. The electromechanical coupling coefficient reaches values between 10 and 15 % and is not high enough for matching of the channel filter dispensing with a coil.

Disadvantages of this concept are the large number of resonators (12 vibrators), the volume of the filter (filter and matching coils 30 cm^3), and difficulties with the spreading size of the welding points (accuracy needed better than 10 μ).

Our concept with torsional resonators and a carrier frequency at 200 kHz is based on 10 years experience in the production of SSB filters[10]. The expense of this concept has been largely reduced by using finite poles, piezoelectric transducers and a new supporting technique [11,12].

A signalling filter and a channel filter are shown in Fig. 5. The signalling filter FZ 23 T consists of 3 torsional resonators (4 mm diameter and 7 mm long) which are interconnected by 2 wires for a weak coupling on the face of the resonators and 2 supporting wires. The channel filter FE 21 T consists of 6 torsional resonators and coupling wires between adjacent and non-adjacent resonators materialising the finite poles.

The composite longitudinal vibrating transducers are coupled to the torsional resonators at opposite points on the circumference. $\lambda/4$ (6,3 mm) was chosen for the length of the coupling wires which have a diameter dependent on the desired coupling coefficient between 90 μ and 180 μ.

The supporting wires are fixed in a node for the torsional mode and point welded to a comb shaped metal sheet which is fixed on a base plate. The transducers are fixed and connected to contact wires for input and output. The transducers reach a coupling coefficient greater than 20 %. Therefore, no coil is needed for the matching of the filter. Because there is no selection by an electric circuit in the filter, spurious modes of the filter must be very weak. Measurements confirm the expected spurious mode suppression.

The volume of the channel filter is not more than 6 cm^3. Some results about the temperature and time stability are given later.

Filter Synthesis

The calculation of the filters was based on methods of the electrical network theory and some new transformations [13]. The different steps in filter design are indicated in Figs. 6 and 7. The desired specifications can be satisfied by a symmetric approximation. Therefore a bandpass-lowpass transformation could be performed and all further calculations shall be limited to lowpass configurations. The approximated function can contain finite poles or allpass parts.

Beginning with a chain matrix $\underline{A}$ of the lowpass approximation (Fig. 6), the poles at infinity are removed in a first step. In the chosen example the poles are represented by the capacitor in parallel with the input and the coil in series with the output. In a second step the remaining chain matrix $\underline{A}'$ can then be replaced by the equivalent circuit consisting of a twoport with the chain matrix $\underline{A}''$ which is in parallel-series connection with a transformer. This equivalence is valid for any reactance twoport [14,15]. The transformer ratio ü should be determined in such a way that the degree of the denominator polynomial at the chain matrix $\underline{A}''$ is reduced with a factor 2.

For the chain matrix $\underline{A}''$ the steps removing poles at infinity and parallel-series connection with a transducer can be repeated till the denominator polynomial of the remaining chain matrix reaches a constant value. Finally, poles at infinity should be removed. The result is a network containing capacitors, coils and transformers.

The circuit is canonical in reactances. In a third step the transformers should be replaced by couplers

$$\underline{A}_K = \begin{pmatrix} 0 & \pm j\,\frac{d}{K} \\ \pm j\,\frac{K}{d} & 0 \end{pmatrix}, \quad (2)$$

which represents a $(2n-1)\cdot\lambda/4$ transmission line in a normalized form as a narrow band approximation. The value d is the damping of the first circuit at the input and K is the coupling coefficient between two circuits.

For the third step in Fig. 6 two properties of the couplers are needed:

1. A chain connection of two equal couplers is equivalent to a transformer.

2. A coupler is a dual transforming network[13].

The transformer in Fig. 6 can be replaced by three couplers wherever the coupler at the output is not interesting for the transmission properties.

With further steps in the described manner the synthesis of the whole network can be performed. Finally, the lowpass must be transformed into a bandpass which represent a mechanical filter as a narrow band approximation. All values, resonance frequencies f_i, dampings d_i and coupling coefficients K_{iK} for the filter are now known. The special task for the designer of mechanical filters is now to calculate the dimensions of the filter parts.

The resonance frequencies for example determine the length of the torsional resonators. The dampings of the input and output circuits should be realized by electrical resistors which are transformed by the transducers tuned to a frequency a little below the frequency of the resonators. The damping of the internal resonators should be very low (Fig. 3). The coupling coefficients determine the ratio of the diameter of the wires and the resonators. Figure 7 shows additional possibilities for changing the coupling coefficient. The coupling with wires on the top of the resonators is weak and grows with the distance from the centre of the resonators. The dashed wire shows a negative coupling - for example for an allpass circuit. Coupling wires on the circumference at the end of the resonators give a stronger coupling which is initially reduced slowly with a fixing point nearer to the middle of the resonators and disappears in the middle of the resonators (supporting wires).

The influence of tolerances on the length of the coupling wires between two resonators depends largely on their length. The merit of a coupler a quarter wave length long is the small influence on the centre frequency and no influence on the bandwidth. Shorter coupling rods, e.g. for the 48 kHz bending filter, cause very high sensitivity at the upper edge of the band (f_1).

Material Problems

The temperature and time stability achieved with mechanical resonators (Fig. 3) is the result of an intensive material research. Special NiFe alloys with small contents of Be (Thermelast 5429) or Ti and Al (Thermelast 4002 or NiSpanC) can be hardened by annealing treatment. The important physical properties of the two alloys are shown in Fig. 8 as a function of the annealing temperature T_A. The temperature change of the resonator frequency Δf_T of such materials is highly dependent on the annealing temperature T_A. $|\Delta f|$ should be the total frequency change in the temperature range from 10°C to 50°C. The minimum of $|\Delta f|$ coincides as a cultivated effect with a maximum in resonator Q. Unfortunately the resonator frequency f_T depends very highly on the annealing temperature T_A, too. For our special purpose the alloy Thermelast 5429 should be prefered due to its broader compensation range and its lower dependence of the annealing temperature T_A[16].

The specifications for the stability of transducers are lower by more than a factor 10. For the shown composite transducers which are assembled from two metal pieces and a ceramic disk by soldering (Fig. 5) a resonance frequency of 200,8 kHz, a capacitance of 200 pF, a coupling coefficient of 21 %, a Q of 2000, a temperature change of 100 Hz in the temperature range 10°C to 50°C, and an aging rate less than 30 Hz/decade has been measured.

The coupling wires are grid wires made of a NiFeMo alloy, which has a tolerance in diameter of 0,5 µ in a batch.

Production Technique

The production process of the filters can be divided in three main parts.

1. Manufacturing and tuning of single resonators and transducers.
2. Point welding of the filter structure.
3. Fitting into a case.

Starting point are drawn rods, from which the resonators and parts of the transducers are manufactured. The tolerances of the dimensions correspond to normal grinding processes (10 µ). As a consequence the spreading of the frequency of untuned resonators and transducers is in the order of ± 3 ‰ (effect of the length ± 1,5 ‰ and effect of spreading in the velocity ± 1,5 ‰). The tuning process must reduce this spreading to ± 2 . 10^{-5} for the resonators and ± 2 . 10^{-4} for the transducers. There is only one tuning frequency for the resonators and one more for the transducers. Tuning is effected by grinding the top of the torsional resonators near the edge of the resonator and in the length of the transducer. This process is performed automatically in a special tuning machine.

For assembling, the resonators, transducer and wires are aligned in a precise gauge and then point welded automatically with controlled welding power. After this process the filter can be measured and needs no adjustment.

Besides the tuning of the resonators, the point welding process has the greatest influence on the transmission characteristic. Calculations with the filter model (Monte Carlo method with 5000 filters), show that for the internal resonators and $\lambda/4$ couplers tolerances smaller than 20 Hz or 1 % respectively must be obtained. The transducers and the $3\cdot\lambda/4$, $5\cdot\lambda/4$ and $7\cdot\lambda/4$ couplers are less sensitive with a factor between 2 and 20. The detuning of resonators by point welding is small enough. The accuracy for the spacings of resonators, the positions of the wires and the welding points can then be determined by reference to Fig. 7.

Fitting into the case is normal mounting work without special difficulties.

Prototypes of the filters are shown in Fig. 9. The coupling wires are not quite straight (an effect of the welding notch) but there is no disturbing influence on the transmission properties of the filters[17].

Our experience with SSB filters shows that channel and signalling filters of the described type could be produced also for a specification of 1/20 CCITT with few rejects.

References

(1) R.A. Johnson and R.Y. Teske,
IEEE Trans. on Sonics and Ultrasonics 13(1966)2,
41 - 48.

(2) Kokusai Electric Co., Ltd.,
Technical Bulletin 6220.

(3) F. Künemund and K. Traub,
Frequenz 18(1964), 277 - 280.

(4) M. Börner,
NFT 19(1960), 34 - 37.

(5) R.A. Sykes and W.D. Beaver,
Proc. 20th Annual Symposium on Frequency Control
(1966), 288 - 308.

(6) H. Kopp,
VDE-Fachberichte 26(1970), 144 - 164.

(7) R.J. Byrne,
Proc. 24th Annual Symposium on Frequency Control
(1970), 84 - 92.

(8) R.A. Johnson, M. Börner and M. Konno,
to be published in IEEE Trans. Sonics and
Ultrasonics.

(9) H. Albsmeier,
Frequenz 25(1971), 74 - 79.

(10) M. Börner, E. Dürre, H. Schüßler,
Telefunken-Ztg. 36(1963), 277 - 280.

(11) H. Schüßler,
Telefunken-Ztg. 39(1966), 429 - 439.

(12) M. Börner,
Radio and Electronic Engineer 29(1965), 173 - 184.

(13) B. Kohlhammer and H. Schüßler,
Wiss. Ber. AEG-TELEFUNKEN 41(1968), 150 - 159.

(14) B. Kohlhammer,
Wiss. Ber. AEG-TELEFUNKEN 43(1970), 170 - 177.

(15) B. Kohlhammer,
Proc. 21th Electronic Components Conference (1971)

(16) H. Albert,
Feinwerktechnik 72(1968), 244 - 248.

(17) H. Schüßler,
ETZ-A 91(1970), 724.

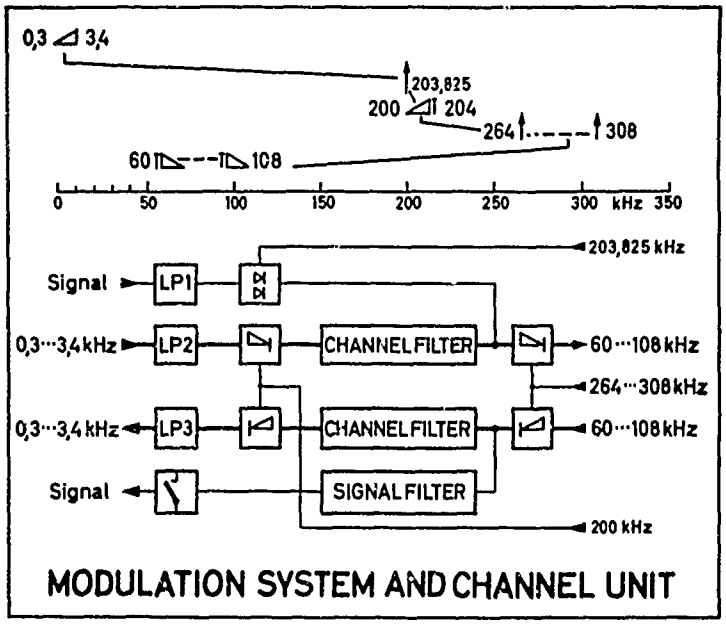


FIGURE 1: Multiplex Plan and Channel Unit for a Premodulation System with Mechanical Filters.

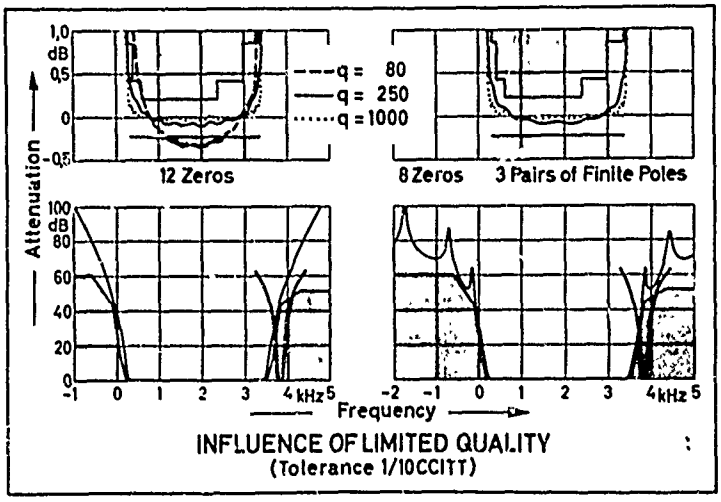


FIGURE 2: Filter Charecteristics of a Polynomial Filter and a Filter with 3 Pairs of Finite Poles.
Normalized Quality Factor q as a Parameter (Tolerance 1/10 CCITT).

FILTER Type	f_0 kHz	Q 10^3	TK(f) $10^{-6}/°C$	$\Delta f/f$ 10^{-6}
LC	24	>2 (0,2-0,5)	<8 (10-50)	<200 (200-500)
BENDING	48	>4 (6-15)	<4 (0,5-2)	<100 (30-100)
TORSION	200	>15 (15-30)	<1 (0,5-1)	<25 (10-30)
MONOLITHIC	8140	>600 (100-200)	<0,025 (0,02-0,1)	<0,5 (0,5-10)

FIGURE 3: Quality Factor, Temperature Coefficient and Aging of Resonance Circuits needed for Channel Filters at different Premodulation Frequencies. Reached Values in Brackets.

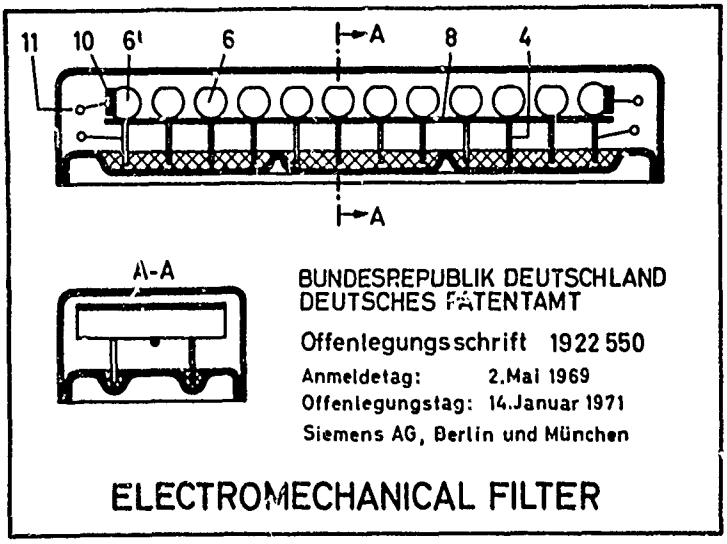


FIGURE 4: Mechanical Filter with Bending Vibrators.

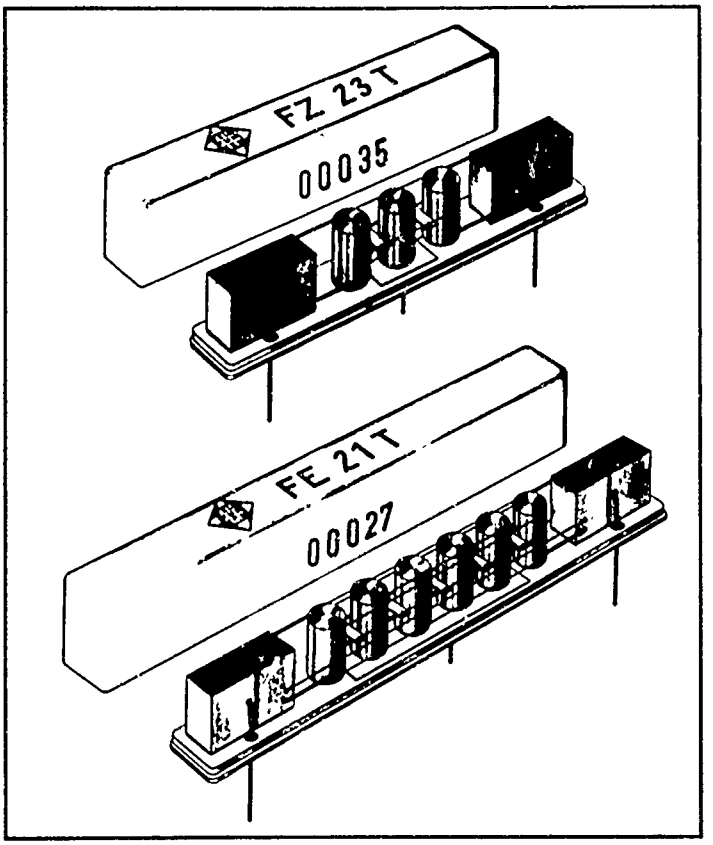


FIGURE 5: Mechanical Signalling and Channel Filter with Torsional Resonators.

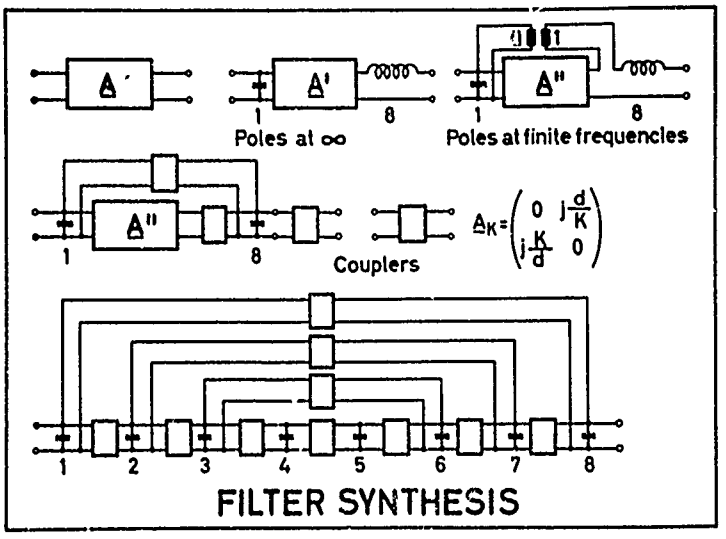


FIGURE 6: Conception of Filter Synthesis for Channel Filters with Finite Poles.

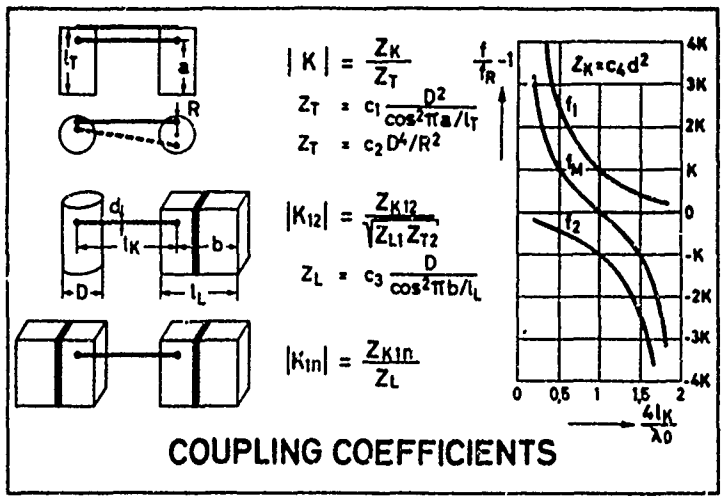


FIGURE 7: Calculation of Coupling Coefficients for Torsional and Longitudinal Resonators.

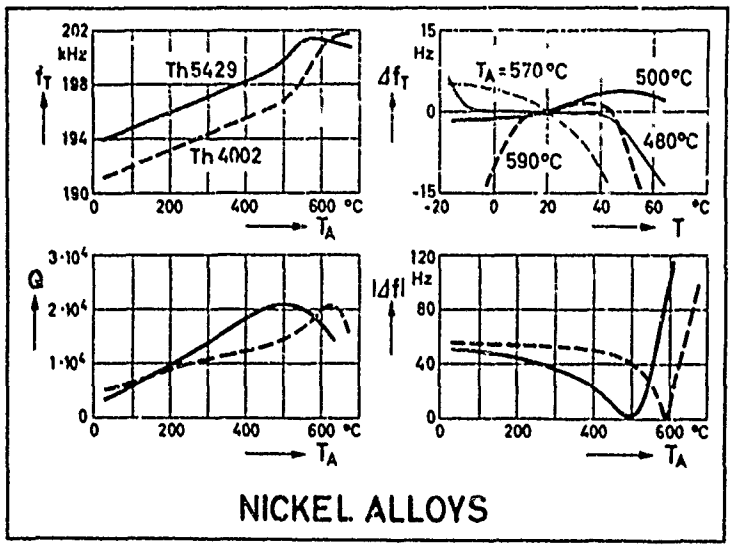


FIGURE 8: Physical Constants of NiFe Alloys for Torsional Resonators in Dependence of the Annealing Temperature T_A.

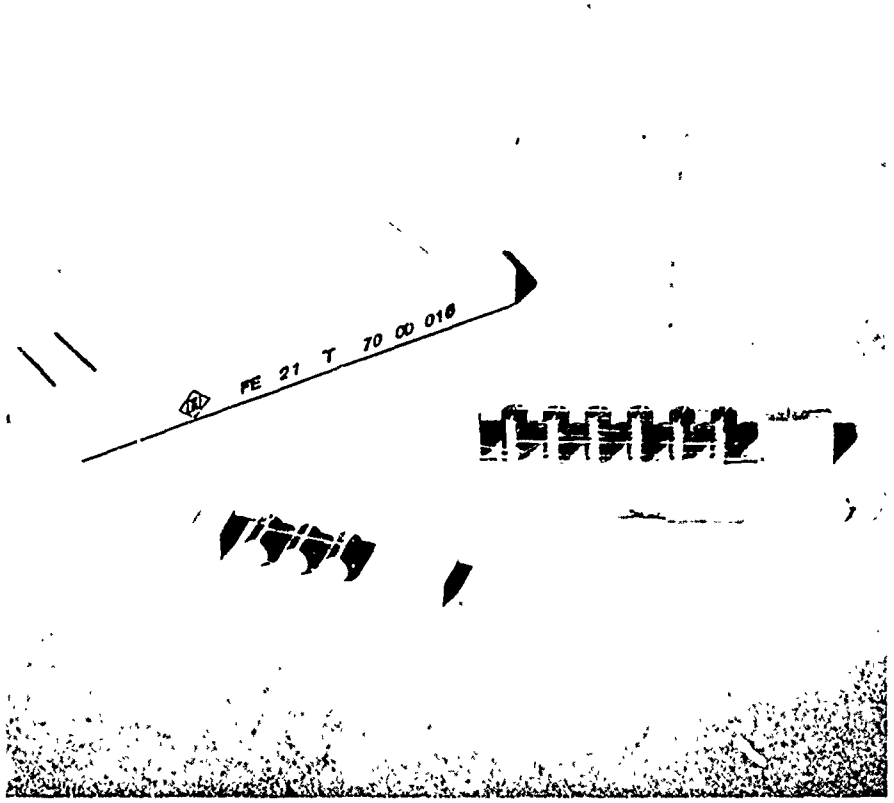


FIGURE 9: Photograph of a Signalling and a Channel Filter.

SEMI - MONOLITHIC QUARTZ CRYSTAL FILTERS

A N D

MONOLITHIC QUARTZ FILTERS

Louis Bidart
Compagnie d'Electronique et de Piezo-Electricité
Sartrouville, 78 - FRANCE

Summary

The purpose of this lecture is to show the evolution, in France, of the conception of a modern quartz filter.

The main technical principles are considered to be well known and the aim of this paper is the discussion of the results and the comparison of various procedures, which will lead to new technics.

Introduction

Up to nineteen sixty, the concepts of quartz filters was conventional and all the devices had an handicraft character. The increasing complexity of the specifications of telecommunications and the wide-cpread use of integrated circuits led to an important transformation and led the manufacturers to look for new solutions, corresponding to modern devices. Fiability and miniaturization, which are the main desired caracteristics of modern electronic equipment, are brought together in monolithic designs.

Monolithic or semi-monolithic filters come from the same synthesis philosophy, and the differences in the definition or in the features, must hide the real analogy that binds the two technics. Finally, the analysis of the experimental results will allow us to bring out the analogy existing between the two designs, even though future devices may lead to diverging developments.

Definitions

At the moment, three kinds of quartz filters exist :

- the conventional quartz crystal filter ;
- the semi-monolithic quartz crystal filter ;
- the monolithic quartz filter.

The conventional quartz crystal filters, which are constituted by discrete components, will continue to exist in ladder structures or by employing the so-called Jauman configuration.

In principle, there is only one definition for a monolithic filter ; that is to say a design which realizes the total integration of all of the components on the same crystal plate. The addition of a discrete component or the seperation of the substrate in several parts, introduces an hybrid feature, which is at variance with the monolithic idea. But it is obvious that there are several degrees in the integration of the components of a network and, even if the definition is not exactly respected, a monolithic style subsists.

From this, one has following definitions :

- The semi-monolithic filter is the result of a chained combination of elementary monolithic filters with, or without, the introduction of linking capacitors and coils.
- The monolithic filter is the result of the complete integration of a network on the same substrate.

In the two devices, one may foresee, inside or outside, two coils or two transformers, which neutralize the stray capacitances at the inputs of the filter. These components do not belong to network and are only used to improve the matching of the filter by the user.

Semi-monolithic filters

They are designed by the cascading of n two-pole resonators. The degree of the filter is even and may attain ten poles. The field of application covers the range 2 to 250 MHz, with, for the AT cut :

- from 2 to 36 MHz in the fondamental mode ;
- " 10 to 108 MHz in the 3rd. overtone ;
- " 10 to 180 MHz in the 5th. overtone ;
- " 15 to 250 MHz in the 7th. overtone.

The admissible bandwiths remain less than

- $3{,}5.\ 10^{-3}$ of F_o for the fondamental mode
- $3.\ 10^{-4}$ of F_o for the 3rd. overtone ;
- $1.\ 10^{-4}$ of F_o for the 5th. overtone ;
- $5.\ 10^{-5}$ of F_o for the 7th. overtone.

The computer program

A program for calculations, giving the synthesis and analysis of polynomial filters, taking loss into account, is represented in the synoptic table in figure n° 1. Starting from the system specifications, one looks for a known mathematical function which approximates as closely as possible the desired response. The transfert function for a polynomial filter is described by :

$$\left|1 + t(j\omega)\right|^2 = \frac{m}{1 + f(\omega^2)}$$

Where m is a constant.

The polynomial function f (ω^2) may be chosen in the bibliography, from :

- Butterworth : f $(\omega^2) = \omega^{2n}$
- Chebyshev : f $(\omega^2) = \xi^2 C_n^2(\omega)$

One may also include Bessel, Gauss, Legendre, Rakovitch, etc... The synthesis procedure which consists of Darlington's insertion loss method, is represented in figure n° 2. For a purely reactive network one has :

$$\left|\rho(j\omega)\right|^2 + \left|t(j\omega)\right|^2 = 1$$

Where ρ is the reflexion coefficient.

1°) Input parameters : the data at the input are as follows :

. Filter : polynomial type
. Number of poles : n
. Ripple (if necessary) : λ
. F_o - Center frequency
. B.W - Bandwith
. L - inductance of the resonators.

2°) Calculations : the polynomial function, which characterizes the transmission function, is established. The loss may be introduced in two different manners :

- shifting the poles and generating a transmission function with loss ;
- taking loss into account during the analysis of the band-pass filter and its optimisation.

From a low-pass normalized filter, the application of the Smythe (1) transformation to the network, leads to a chain of lattices containing one quartz per arm ; these lattices are coupled together by means of capacitive or inductive lattices.

The existence of this capacitor lattice is not desirable. When there are no problems for the phase linearity and the filter may be considered as a narrow band-pass filter, the capacitive lattice may be reduced to a single capacitor.

An optimisation subroutine coupled to an analysis programme allows one to modify the network parameters until the desired response shape is obtained.

Then, to each quartz lattice section, corresponds a two-pole monolithic filter. Then, another subroutine computes the masks, taking into account the reversibility principle.

3°) Reversibility principle : The function of a two-pole resonator is represented by (2)

$$F_o \cdot BW \cdot \alpha \left(d + 2r\frac{a,b}{\xi}\right)^2 - e^{-\xi d} = 0$$

with the notations of figure n° 3.

The constants a, b and α are partly linked to the geometry of the electrodes. The variables are : the plate back and the dimensions of the electrodes.

A physical principle tells us that for each frequency, there is an optimum relation between the mask dimensions and the plateback, leading to a minimum for the level of unwanted modes.

Therefore, for each filter problem, only one mask is possible. In fact, as soon as the coincidence of spurious responses is avoided, the use of a mask may be generalised, when the stopband attenuation meets the specifications. Under these conditions, an optimized mask given for a special problem, may be used for various filters, in a wide range of frequency.

The reversibility principle is such that the synthesis of a filter gives a mask, and this mask may be used in other applications, by modifying the mass loading or the orientation of the substrate.

4°) Non-coincidence of spurious responses : A two-pole monolithic section is, by definition, a multiple-bandpass filter. As in the case of the classical spurious responses, an unwanted bandpass exists. The level of spurious responses may be between 6 and 50 db. for a section.

The cascading of two identical sections may give a spectrum of unwanted frequencies between 12 and 100 db. But, if we realize two sections from the same mask with differences in the mass loading or in the orientation of the substrate, there will not be coincidence for spurious responses. The cascading of these two sections will give an improved stopband attenuation, that can reach 100 db., even from resonators individually showing values of about 15 db.

The reversibility principle for the masks, in addition to the precautions taken to avoid coincidence of spurious responses, is one of the industrial advantages of this device. We also see that the stopband attenuation of a semi-monolithic filter, can easily reach 100 db. and that the suppression of spurious responses is not a difficult problem.

5°) Output of the program : with all the parameters for the technical realization, the output of the program gives the amplitude response, the phase response and the reflective attenuation. (3)

Results

Figure n° 4 represents a typical realization of a semi-monolithic filter, for 25 kHz channels. This new construction has definitively surpassed the traditional devices. In general, we may say that all the filters which equipped AM or FM or data transmission receivers, for

50, 25, 20 or 12,5 kHz channels, are now manufactured with semi-monolithic filters, in the range of frequencies between 10 and 30 MHz.

Figure n° 5 shows the characteristics for a filter with 10 poles for a single sideband receiver. This is an application, in 3rd. overtone, which is now developed.

A 175 MHz filter is represented in figure n° 6. It is the result of studies begun in 1970, to obtain quartz filters between 100 MHz and 200 MHz, where the design of classical differential transformers is particularly difficult.

Discussion

The advantages of semi-monolithic filters clearly appear, as soon as we compare them to classical networks. The semi-monolithic conception is characterized by :

1°) the suppression of the differential transformers ;
2°) the integration of (at least) 2 poles on the same crystal plate.

. Miniaturization - the volume obtained is three times smaller than for a conventional device.
. Fiability - the fiability is assured by the reduction of discrete components, under the climatical and mechanical environment.
. Electrical characteristics - the suppression of differential transformers gives better stability to the output versus temperature, and for the same reason, a high stop-band attenuation (around 100 dB.).

o-O-o

To sum up, we may say that this modern conception of quartz filters, has enough technical advantages to surpass the traditional design. The device is characterized by ease of application :

. by the use of conventional equipment ;
. by the use in a wide range of frequency ;
. by the reversibility of the masks, which is the main problem for the manufacturer.

Monolithic filters

They are obtained by the integration of n poles on the same substrate, excluding any other component. The number n may be increased to 4 and 8, without theoretical limitations, but technological problems may still exist.

The computer program

A synoptic table in figure n° 7 shows the synthesis of monolithic filters. The mathematical principles are identical to those used for the synthesis of semi-monolithic filters, but the program is slightly different and the calculations do not take loss into account.

1°) Input parameters : they include :
. the transmission function ;
. the center frequency ;
. the bandwith ;
. the ripple (if necessary) ;
. the dimensions of the electrodes ;
. the resonating frequency of the substrate.

2°) Calculations : the transmission function gives a low pass filter. Then, the coupling coefficients kij are found and allow the calculation of the distances dij between the electrodes. The mask is obtained and we have all the elements to realize the filter.

On the other hand, the low pass filter is transformed into a passband filter, as shown on figure n° 4. We must notice that the three diagrams are equivalent, even if the section with several quartz per arm, is the only one which represents the physical phenomenon.

Consequently, any of these diagrams may be analysed and the results compared, taking into account the requirements and the precision of the critical frequencies of the device. If it is necessary to introduce loss, the results of the analysis is fed to the input of the synthesis, to compute all the parameters of the mask.

Inversely, a mask which gives wrong results, may be studied in its design, to compute a new filter, using the plate back as a variable. Again, we find some features of the reversibility principle that we discussed for semi-monolithic filters. The function giving the distances is represented by (4) :

$$F_o \cdot BW \cdot \gamma \cdot K_{ij} \left(d_{ij} + 2r + \frac{a \cdot b}{\xi} \right)^2 - e^{-\xi d_{ij}} = 0$$

where γ, a, b are related to the orientation, the geometry and the number of electrodes.

Experimental results : the characteristics of a 4-Pole monolithic filter and of a set of 2 x 4 poles at 8,5 MHz are described in figure n° 7. Figure n° 8 represents an 8-Pole filter, at 10,7 MHz.

Electrical characteristics :

The low insertion loss allows the calculations to be carried out for a filter without loss, even for narrow pass bands. The ripple, which increases with the number of poles, remains negligible. The selectivity fits the specification transfert function and the number of poles. The stopband attenuation is limited by the spurious level, which is due to the shape of the electrodes or to the reflexions at the ends of the crystal plate. In spite of foreseen technical improvements, the unwanted modes often reach the level of 60 db. (independent of the number of poles). This is a weakness of the design, but it is not a handicap for its development.

Sometimes, matching requires auto-transformers, situated outside of the filters, to fit the characteristics asked by the user. We must notice that, as opposed the case of semi-monolithic filters, they can give a larger band to the filter.

Discussion

All the considerations concerning the reduction of the components and its consequences of miniaturization, fiability and stability, are similar to those studied in the case of semi-monolithic filters. The results are noticeable for the bandwith and the insertion loss. The stopband attenuation is limited and the increase of the number of poles has no reaction on the level of spurious responses. But, as for the semi-monolithic filter, the cascading of two monolithic networks with four poles, may give a stopband attenuation larger than 100 db.

Finally, the main problem is the difficult design of the crystal plate, due to its large dimensions. Because of its mechanical weakness, the application is reduced to a bandwith that we may consider from 6 to 13 MHz in the fondamental mode.

Conclusion

The comparison of the two designs that we analysed is possible for the same frequency, for example 10 MHz, and it is obvious that the semi-monolithic filter is more efficient when it is used in a wide range of frequency and its stopband attenuation reaches 100 db., which gives a good protection from jamming. The manufacture of semi-monolithic filters, which is not very different from the classical filters, seems to be easier. The two designs are similar concerning miniaturization.

The fiability should be greater for a monolithic filter, but experiments may invalidate this theoretical principle. The low insertion loss and the smaller ripple are unquestionable advantages, compared to the results obtained from the semi-monolithic filters which have their own superiority over the classical networks. The design of monolithic filters requires a new technology and the manufacture necessitates an important investment. In any case, we believe that the design must be studied in all its details, because, even if the semi-monolithic filters, with their ease of application, seem to fit all the specifications of modern telecommunications, monolithic filters, taking into account technical and economic considerations, are a solution for single sideband filters, for example, or when the contract calls for 50.000 sets with the same specifications.

The difference between the two designs is not very important as soon as we consider that they come from the same synthesis. The integration of a network shows the technical level of the manufacturer. The production, starting from semi-monolithic filters, leading to the design of pure monolithic filters, will result in better quality in the devices and the leadership.

Acknowledgments. Thanks are due to the members of the laboratories of telecommunications of the Army, for their constant interest and for the mathematicians of the Data Center of the THOMSON - C.S.F. Company, for their valuable assistance during the last four years.

References

(1) B.C. Smythe, "The synthesis of Crystal-Capacitor tandem Lattice All-Pole Bandpass Filters as the insertion - Loss basis" 18 th. Annual Symposium on frequency control.

(2) R.A. Sykes and W.D. Beaver, "High frequency monolithic crystal filters with possible application to single frequency and single side band use".

(3) A. Biroir, "Un programme d'analyse et de synthèse des filtres". Onde électrique; Octobre 1969.

(4) W.D. Beaver, "Theory and design principles of the Monolithic Crystal Filter". Lehigh University. 1967.

SEMI-MONOLITHIC QUARTZ CRYSTAL FILTER

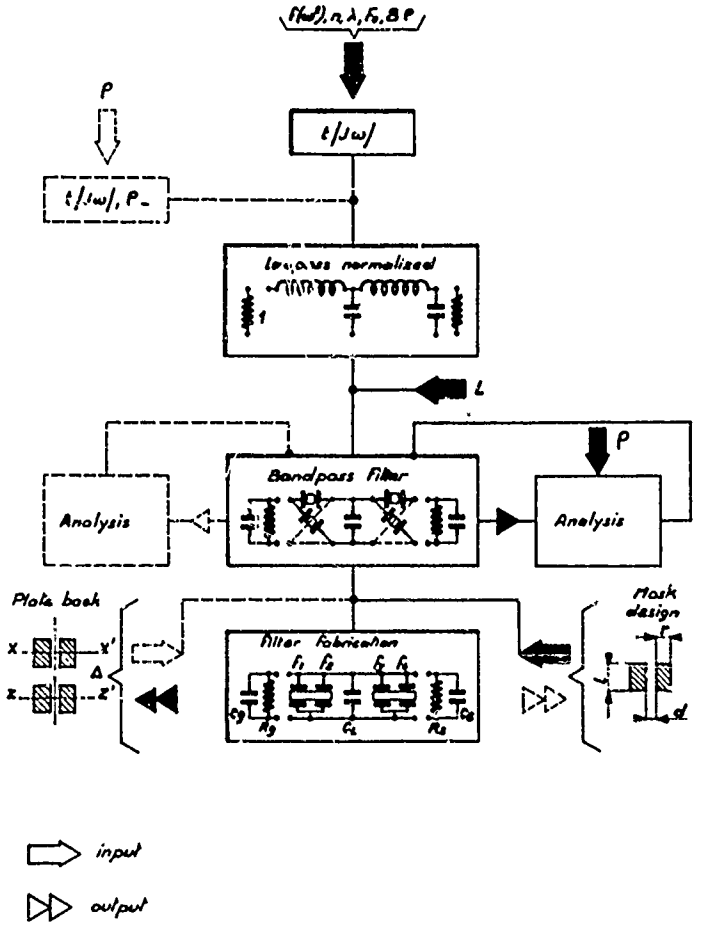
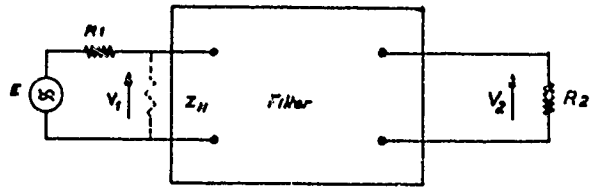


Figure 1

Design procedures for Synthesis programme

INSERTION LOSS TECHNICS



Maximum power : $P_m = \frac{E^2}{4R_1}$

Power in the load : $P_2 = \frac{|V_2|^2}{R_2}$

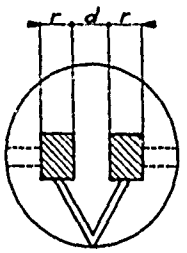
Insertion loss . $|t(j\omega)|^2 = \frac{P_2}{P_m}$

Reflexion coefficient . $\rho = \frac{R_1 - Z_{11}}{R_1 + Z_{11}}$

Figure 2

NON-COINCIDENCE

UNWANTED MODES



F_0 = in KHz

BP = in KHz

r = in mm

d = in mm

$x = \frac{l_2\sqrt{\Delta}}{t}$

With $\Delta = \frac{F_a - F_r}{F_s}$ (Plate back)

F_s = in KHz and $F_s = \frac{167}{t}$

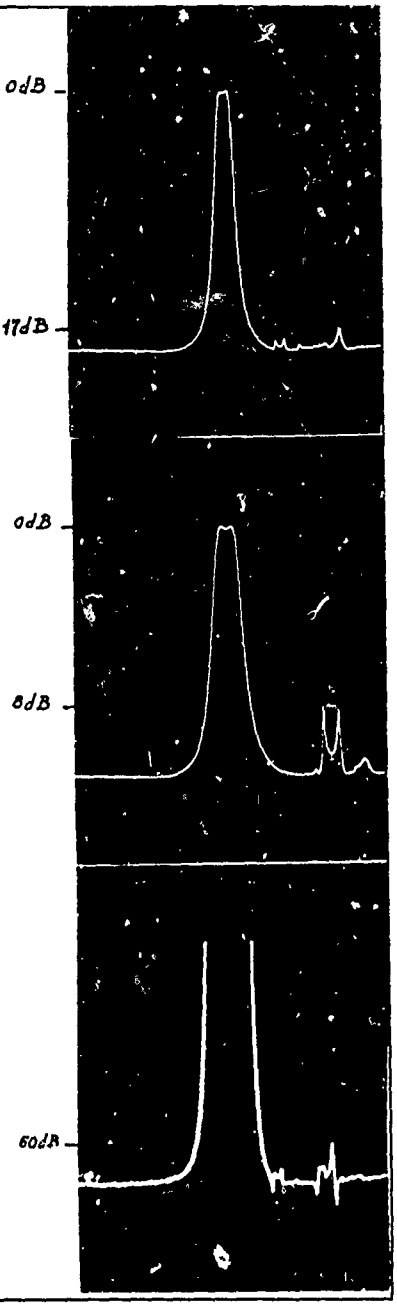
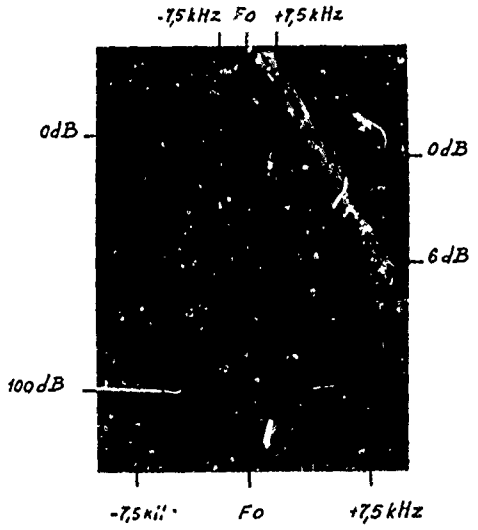


Figure 3

SEMI-MONOLITHIC QUARTZ CRYSTAL FILTER

F_0 = 10,7 MHz (25 KHz CHANNEL)



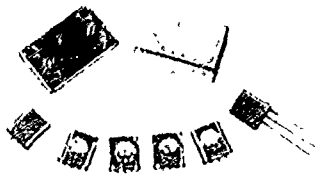
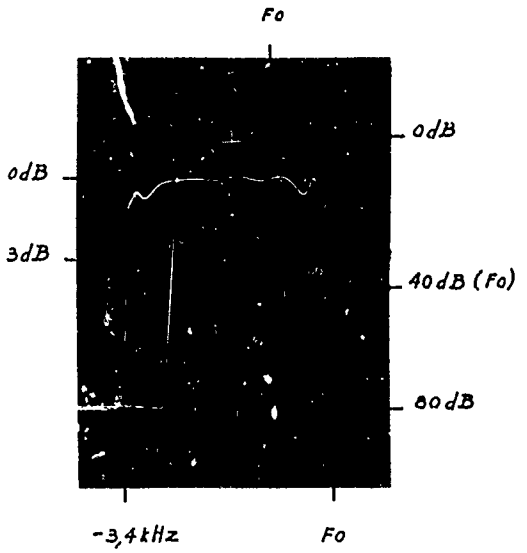


Figure 4

SEMI-MONOLITHIC QUARTZ CRYSTAL FILTER

Single sideband F_o = 12,5 MHz



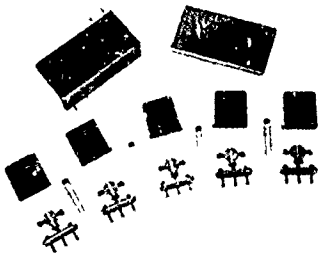
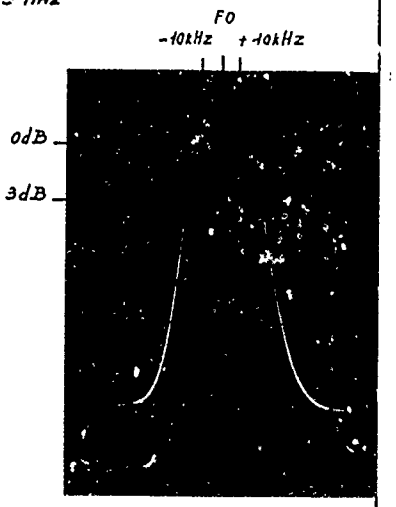
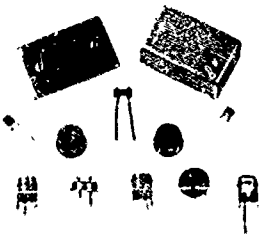


Figure 5

SEMI-MONOLITHIC QUARTZ CRYSTAL FILTER

F_o = 175 MHz





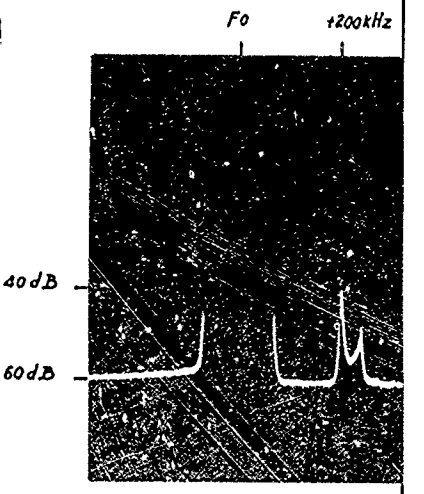
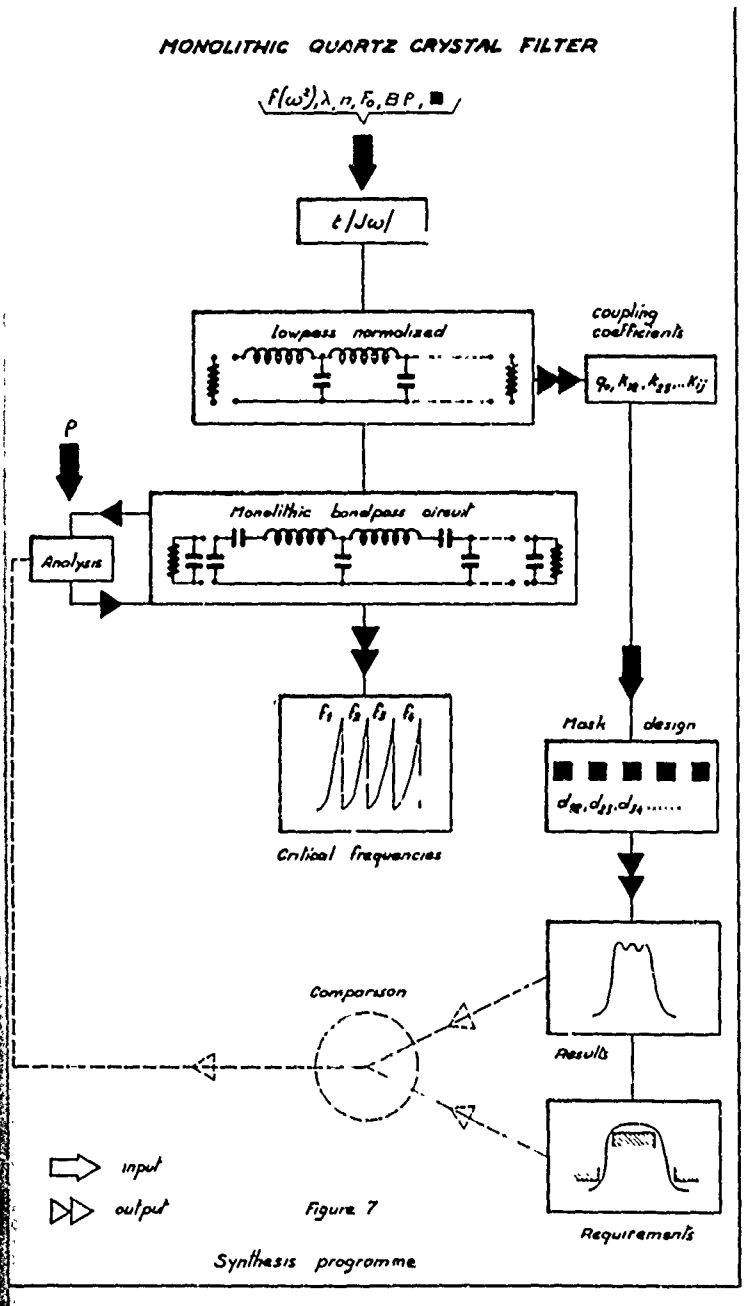
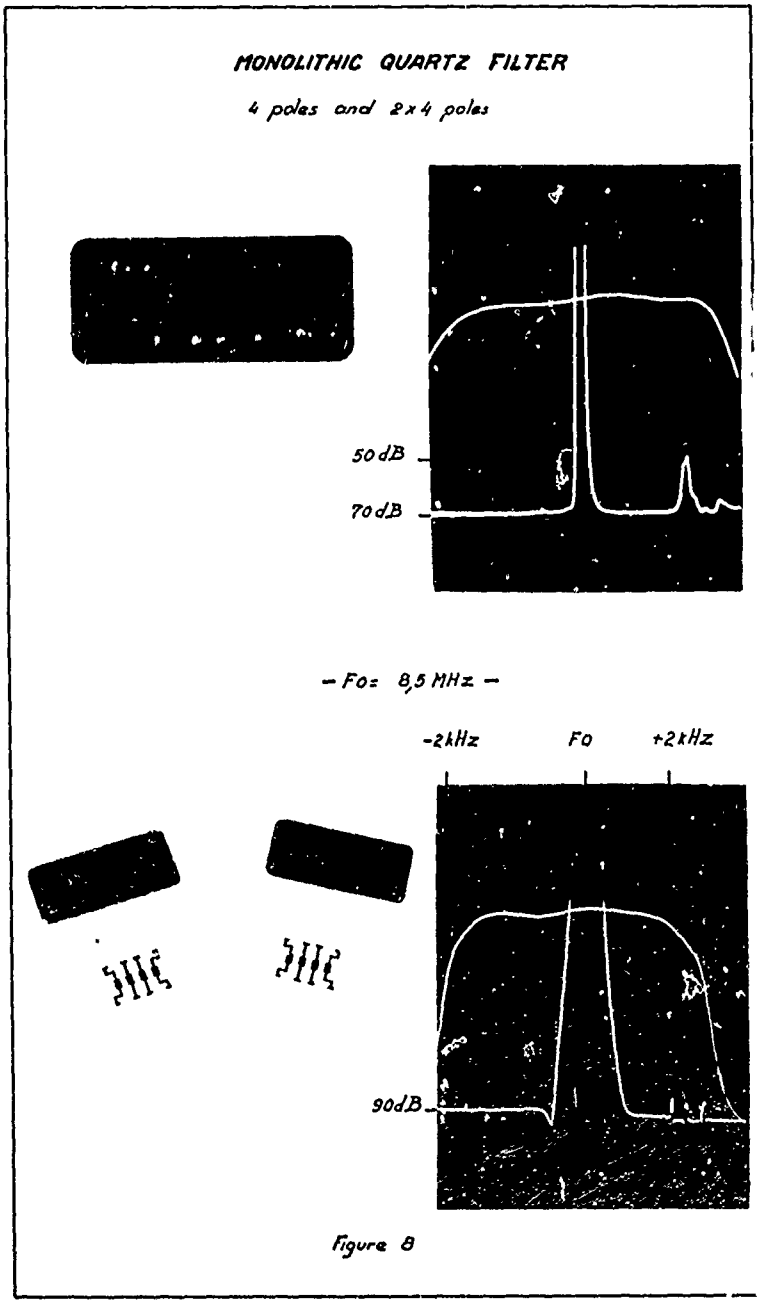


Figure 6

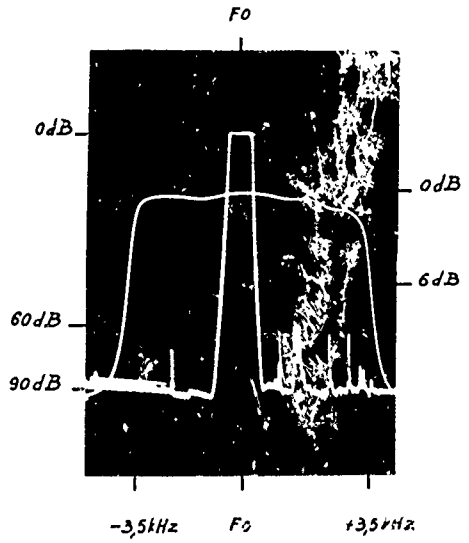

MONOLITHIC QUARTZ CRYSTAL FILTER
f(ω³), λ, n, F₀, BP, ■
t/Jω/
lowpass normalised
coupling coefficients
q₀, k₁₂, k₂₃...kij
P
Monolithic bandpass circuit
Analysis
f₁ f₂ f₃ f₄
Critical frequencies
Mask design
d₁₂, d₂₃, d₃₄.....
Comparison
Results
Requirements
input
output
Figure 7
Synthesis programme


MONOLITHIC QUARTZ FILTER
4 poles and 2x4 poles
50 dB
70 dB
– Fo= 8,5 MHz –
-2kHz
Fo
+2kHz
90dB
Figure 8

MONOLITHIC QUARTZ FILTER

8 poles

- Fo = 10,7 MHz -



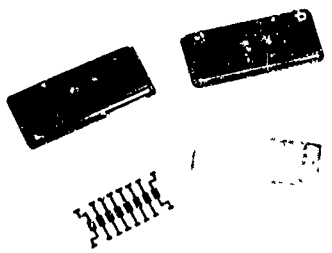


Figure 9

MONOLITHIC CRYSTAL FILTERS FOR FREQUENCY DIVISION MULTIPLEX

by

P. Lloyd
Bell Telephone Laboratories, Inc.
Allentown, Pennsylvania

Abstract

The purpose of this paper is to review recent development work on monolithic crystal filters for a new frequency division multiplex terminal (A6 Channel Bank) to be used for long haul transmission. From a systems standpoint, the new channel bank will duplicate the performance of the existing A5 channel bank but offers considerable savings in both size and cost.

While MCF's are already used in a number of applications, the stringent requirements and complexity of the channel bank filter required the development of a new technology suitable for high volume production and low cost.

Recently, Bell Laboratories established a Process Capability Line for monolithic crystal filters at the Western Electric Plant in North Andover, Mass. The objectives of this line were to transfer this technology to the Western Electric Company and to develop facilities suitable for the evaluation of processes prior to manufacture. A description of these facilities and processes will be given and their design related to the performance objectives of the filter. Statistical results on the performance of devices fabricated will be related to fundamental studies.

Introduction

The largest single application of crystal filters in the Bell System is in voice channel filters for frequency division multiplex. Twelve voice channels are assembled or disassembled in an internationally agreed group between 60 and 108 kHz. The current filter contains four length extensional mode quartz resonators coupled by inductors, capacitors and resistors. A total of 48 different resonator designs are used in a complete A5 channel bank.

Towards the end of 1967, Bell Laboratories began the development of new channel bank (A6) using monolithic channel filters operating in the 8 MHz range. Fig. 1 shows the in-band requirements for these filters and the theoretical attenuation characteristic of an 8 pole, .1 dB ripple Chebyshev filter designed to meet these in-band requirements and also to provide sufficient attenuation in the skirt regions of filter to meet inter-channel crosstalk suppression requirements. The in-band requirements shown in this figure in no way represent what is possible in terms of performance with monolithic crystal filters. Compatibility with existing plant of over one million two-way circuits requires that the in-band requirements remain essentially the same as in earlier channel filter designs.

Considerable cost savings are possible with a monolithic channel filter at 8 MHz; since only one basic design is involved full advantage can be taken of high volume production techniques similar to those widely used for other planar devices. However, in order to realize these advantages an increase in precision of two orders of magnitude must be made over the existing filter. For example, the requirements at the low end of the band call for a ± 25 Hz absolute tolerance, or ± 3 parts in 10^4 at 80 kHz and ± 3 ppm at 8 MHz.

A6 Channel Filter - Introduction of a New Technology

Fig. 2 shows the A6 channel filter in its current form. The device consists of an AT-cut quartz plate with eight pairs of electrodes. Small metallic stripes are deposited between the electrodes, on the top surface of the plate, to provide for coupling adjustment. The plate is suspended from a rectangular annular ceramic mount by thermocompression-bonded leads. The mount provides access to all eight resonators for measurements during processing. The plate is cut in two after mounting and in the final assembly of the filter the two sections of the plate are coupled with a mica capacitor.

Byrne[2] and Miller[3] described the design of this device and the development of associated processing techniques at last years' Frequency Control Symposium. At that time a number of laboratory models of the filter had been constructed and evaluated in an experimental channel bank. Somewhat earlier, namely in the latter part of 1968, it became evident that the techniques needed for the fabrication of the device were such a departure from existing quartz crystal technology that this new technology could best be transferred to the Western Electric Company (Manufacturing Unit of the Bell System) by the establishment of a Process Capability Line (PCL), a concept used successfully in the introduction of integrated circuit technology. While, technical direction of the PCL was through Bell Laboratories it was, nevertheless, a co-operative venture with the Western Electric Company, being staffed by teams consisting of both Western and Bell Labs engineers. The PCL for the A6 channel filter was designed and installed at the Western Electric Plant in N. Andover, Mass. during the last two years. During this period channel filter development has been principally in the area of fabrication techniques and associated facilities for this PCL.

MCF Fabrication in the PCL

Fig. 3 is a block diagram of the filter fabrication steps. With the exception of the quartz growing facility all of the steps shown are located within the PCL. The PCL facilities were designed with a through-put of 60/hr. as an initial objective but no attempt was made to balance the line by duplication of facilities where this objective was unreasonable. The operations for cutting, lapping, grinding and etching the plates involve commercially available machines and with exception of some automation for the sorting the plates, the process is essentially as described by Miller,[3] so in the remainder of this paper a description will be given of the major facilities used for the deposition of the electrode array and the adjustment of the filter. These facilities have

required considerable effort and have been designed and built from the ground up.

Initial Metallization

Fig. 4 summarizes the initial metallization operation. A 6000Å film of Ti-Pd-Au is deposited on the major surfaces of the quartz plate. The Ti-Pd-Au metal system was selected because of its proven use in the thermo-compression bonding of ceramic substrates in the integrated circuit field. Evaporation is from multi-hearth electron gun sources. Plateback and evaporation rate is monitored by quartz crystal monitors. Uniformity of deposition is achieved by careful geometrical design and rotating substrate holders. The facility operates under closed loop control. Random variations in resonator frequencies including random variations in plate thickness and mask geometry are 5 kHz or a 1% control of the plate back. Due to a systematic roll off at the edges of the quartz plate, the resonators have a systematic variation along the plate of about 5 kHz.

Frequency Adjustment

Fig. 5 illustrates the frequency adjustment facility where these systematic and random errors, existing after initial metallization, are reduced to about 12 Hz by further deposition of Au. Eight filters (64 resonators) are loaded at a time into the machine, which is completely automatic. The diagram illustrates the adjustment of resonator No. 5. When this adjustment is complete a shutter is closed and the machine control system causes the device to be moved until resonator number 6 is over the plating aperture. This process continues until all resonators have been adjusted. Now that all resonators on the plate are at approximately the same frequency the inter-resonator coupling can be adjusted.

Laser Trimming

Fig. 6 shows a block diagram of the laser trimming facility. Coupling is adjusted by machining the small stripes deposited between the electrodes.[4] The device is held on an X-Y positioning table and moved under the laser beam. The electrical measurements, the pulsing of the laser and positioning of the crystal are co-ordinated by a tape-programmed controller. Apart from loading and unloading, the machine is completely automatic. Following coupling adjustment one more adjustment of the resonators must be made. This adjustment can be performed by either the frequency plater or alternatively by the laser facility in which case small holes are vaporized in the electrode areas.

Final Adjustment and Sealing

The most critical requirement on the channel filter is that the lower 3 dB point must be held to a frequency accuracy of ± 3 ppm for all causes including aging and temperature variations. For this reason one last adjustment of the filter is saved until the device is mounted on a header.[2,5] Fig. 7 illustrates the machine designed for both the final adjustment of the filter and the cold welding of the cover to the header. The device is loaded into the upper half of a die which is then lowered onto a transfer table. The device is then positioned over a masked evaporation source and is adjusted (typically 50 Hz). At the next table position the upper and lower die halves arrive together for the cold welding operation. The sealed device is then returned to the loading position, removed and another device inserted.

Testing and Data Analysis

With the device now complete, testing of the filter characteristics against requirements is performed with the computer controlled transmission measuring set shown in Fig. 8. While the hardware shown here is relatively simple, considerable sophistication in testing is possible with program control. In fact, this facility has been the prime source of data for the evaluation and control of the process since it is capable of making very accurate measurements of equivalent circuit parameters of devices in-process as well as the measurement of filter characteristics. Data collected from this facility and other information such as orientation angles, plate uniformity etc. has been logged into a time sharing computer for correlation studies. Fig. 9 shows some results for 314 filters fabricated by mid-1970. As can be seen, some systematic errors from the theoretical curve existed at that time - these errors have since been corrected. Fig. 10 shows a plot of the standard deviation expected from the theoretical filter characteristic assuming perfect adjustment of 3 dB point. Values for the variances in frequency adjustment, coupling adjustment and device Q, typical for mid-70, were used. The standard deviation of the 314 units of Fig. 9 is compared with these theoretical results. The error in 3 dB point adjustment was removed for the 314 units by the computer. Again the agreement with the theoretical curve is good except in the frequency range where systematic errors between the average curve and the theoretical curve were apparent in Fig. 9.

The arrows in Fig. 10 indicate the limits of the filter requirements and clearly demonstrate the precision to which the filter can be made.

Conclusion

During the last few years, monolithic crystal filters have been developed and an appropriate process technology has been established for the channel filter used in frequency division multiplex. This work has had considerable impact on analogue communications in the Bell System and is expected to lead to further applications of MCF technology in the future.

Acknowledgment

This work has been a group effort involving a team of Bell Laboratories and Western engineers over several years. The results of their efforts have been reported in this paper.

References

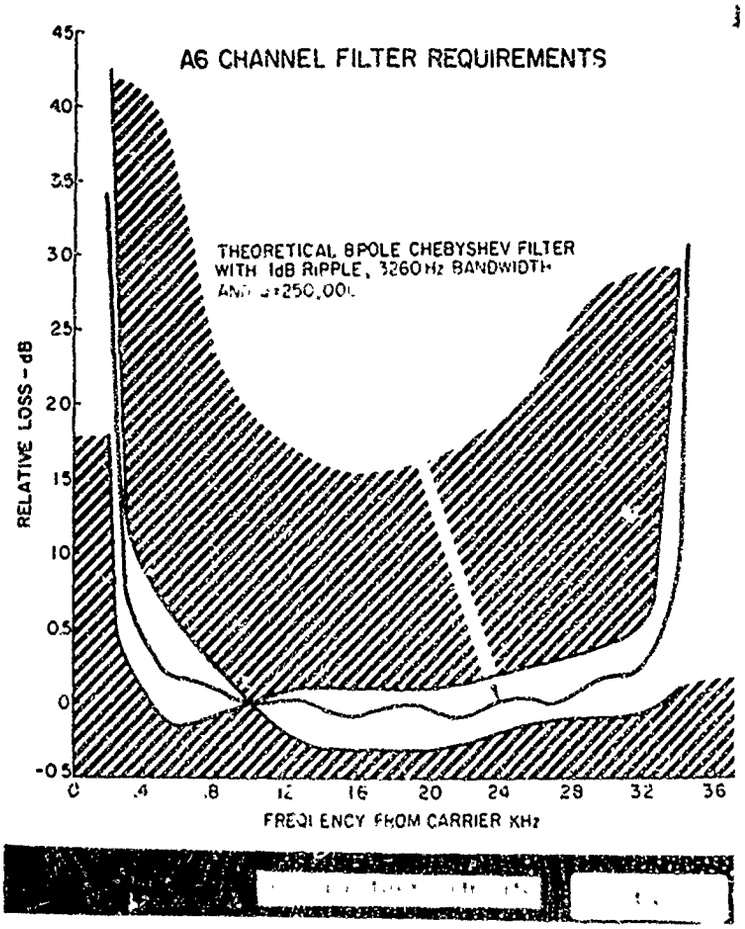
1. G. W. Bleisch, The A6 Channel Bank to be presented at the International Communications Conference, Montreal, June 1971.

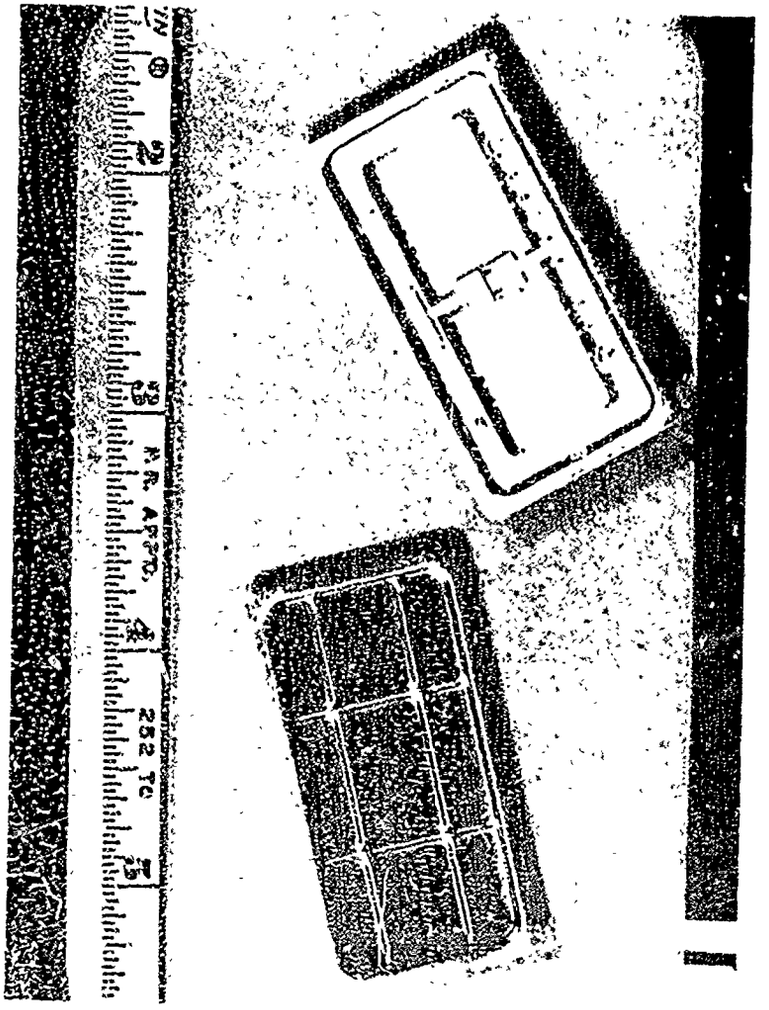
2. R. J. Byrne, Proc. 24th Annual Frequency Control Symposium, April, 1970.

3. A. J. Miller, Proc. 24th Annual Frequency Control Symposium, April, 1970.

4. J. L. Hokanson, Proc. 23rd Annual Frequency Control Symposium, April, 1969

5. R. P. Grenier, Proc. 24th Annual Frequency Control Symposium, April, 1970.


A6 CHANNEL FILTER REQUIREMENTS
THEORETICAL 8POLE CHEBYSHEV FILTER
WITH 1dB RIPPLE, 3260Hz BANDWIDTH
RELATIVE LOSS - dB
45
40
35
30
25
20
15
10
05
0
-05
0
.4
.8
1.2
1.6
2.0
2.4
2.8
3.2
3.6
FREQUENCY FROM CARRIER KHz



FABRICATION STEPS

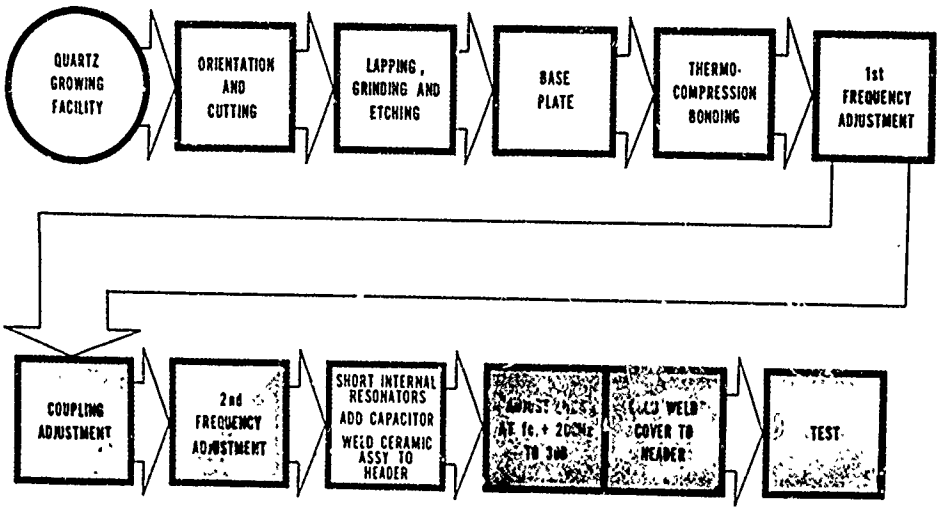
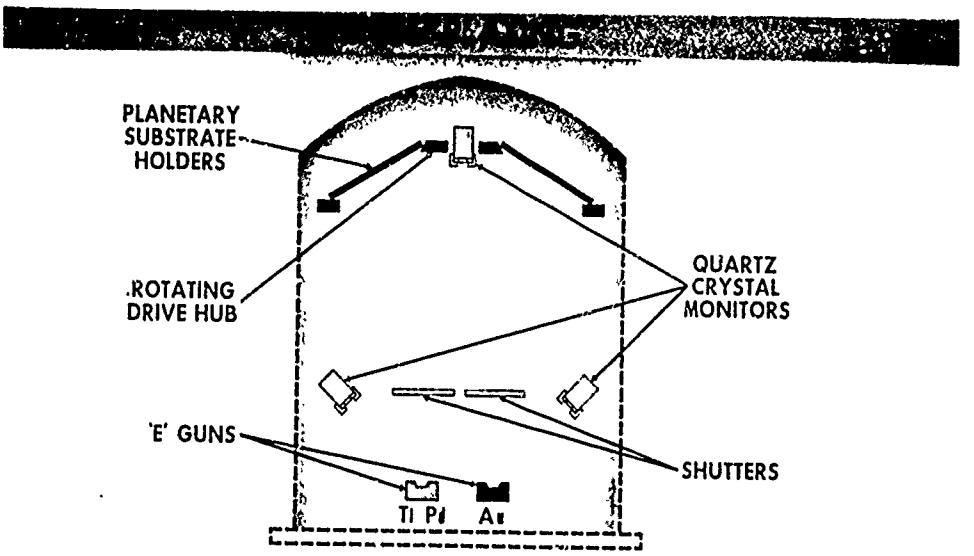


Fig. 3



DEPOSITION	CAPACITY	PLATEBACK CONTROL	FREQUENCY SPREAD (ALL CAUSES)
TITANIUM 2000 Å PALLADIUM 1000 Å GOLD 3000 Å	60 CRYSTALS/HOUR (2 SIDES) 30 CERAMICS/½ HOUR	CHANNEL FILTERS 227KHz +0/-4KHz CARRIER SELECTION FILTERS 242KHz +0/-4KHz 1% OF 12000Å	RESONATORS 1-8 4.4KHz AVE. RESONATORS 4-5 4.6KHz AVE.

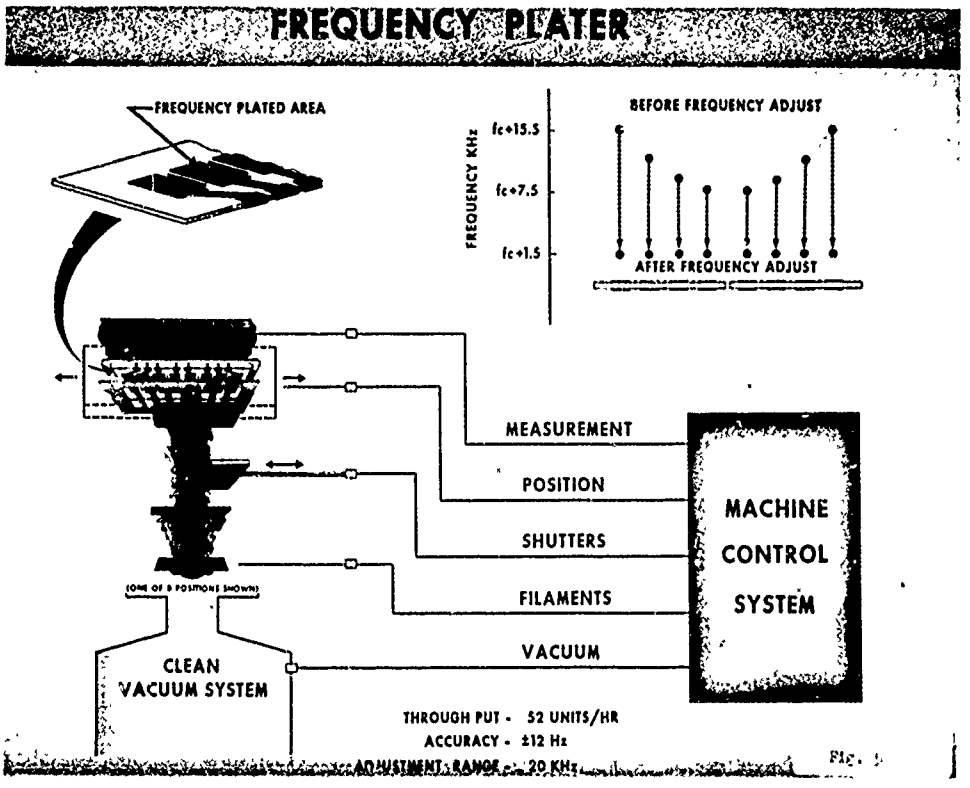
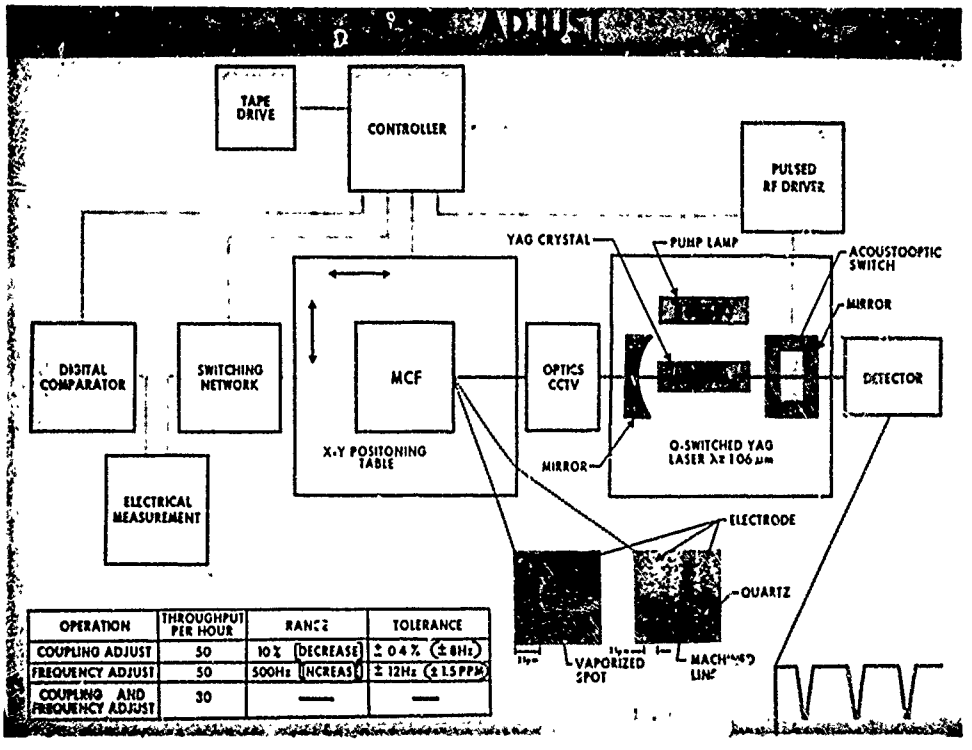
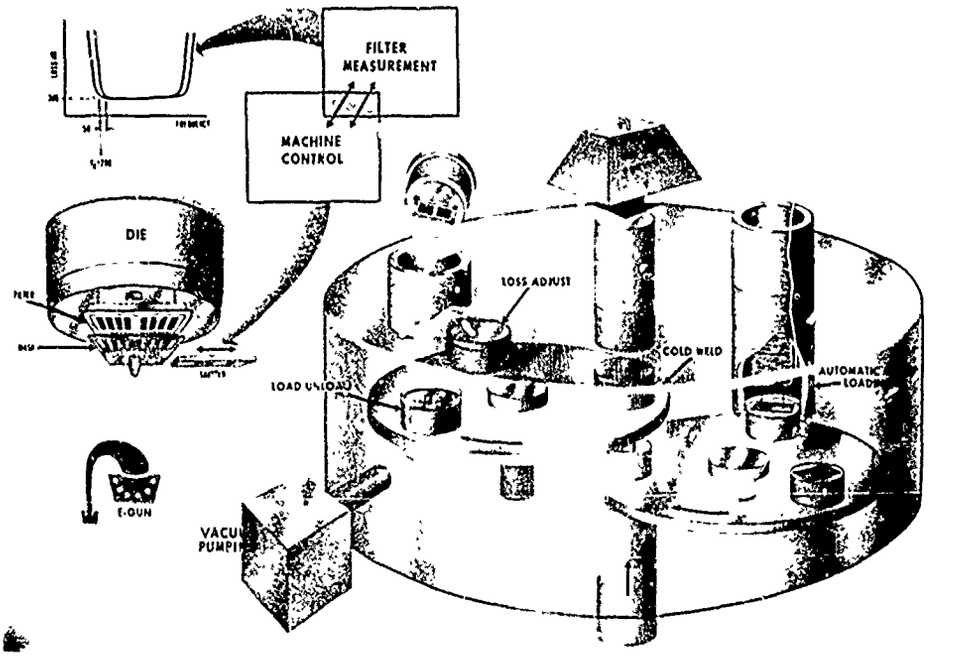


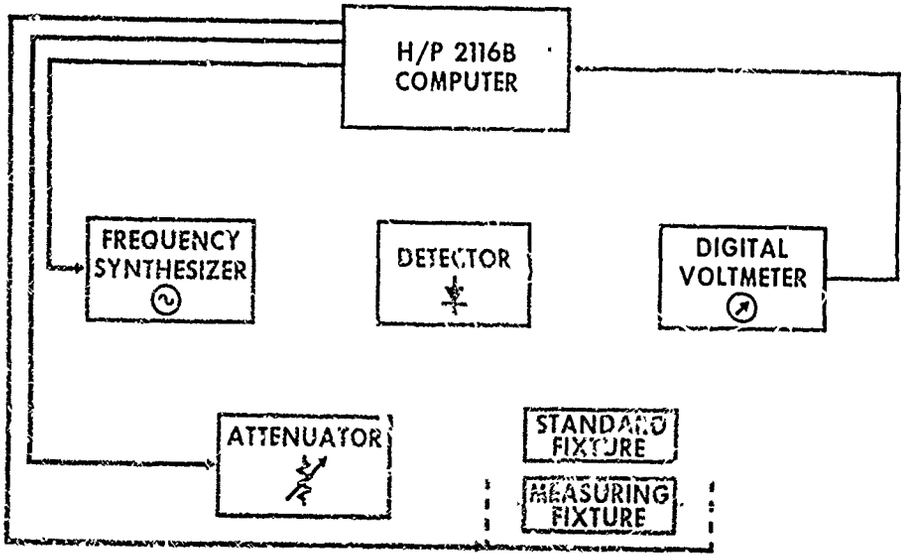
Fig. 1

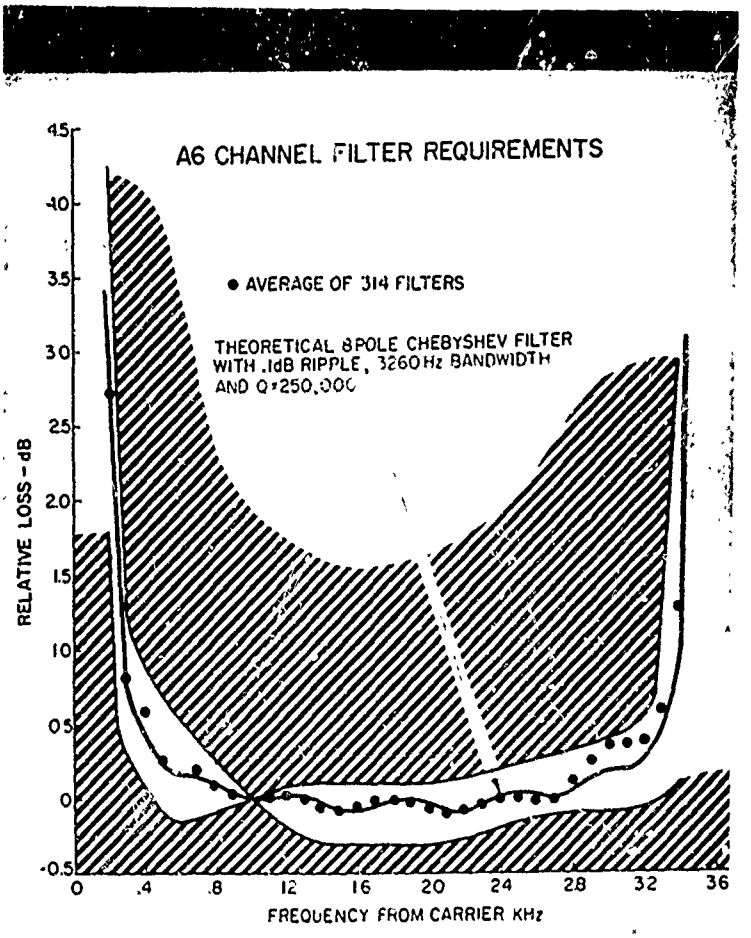


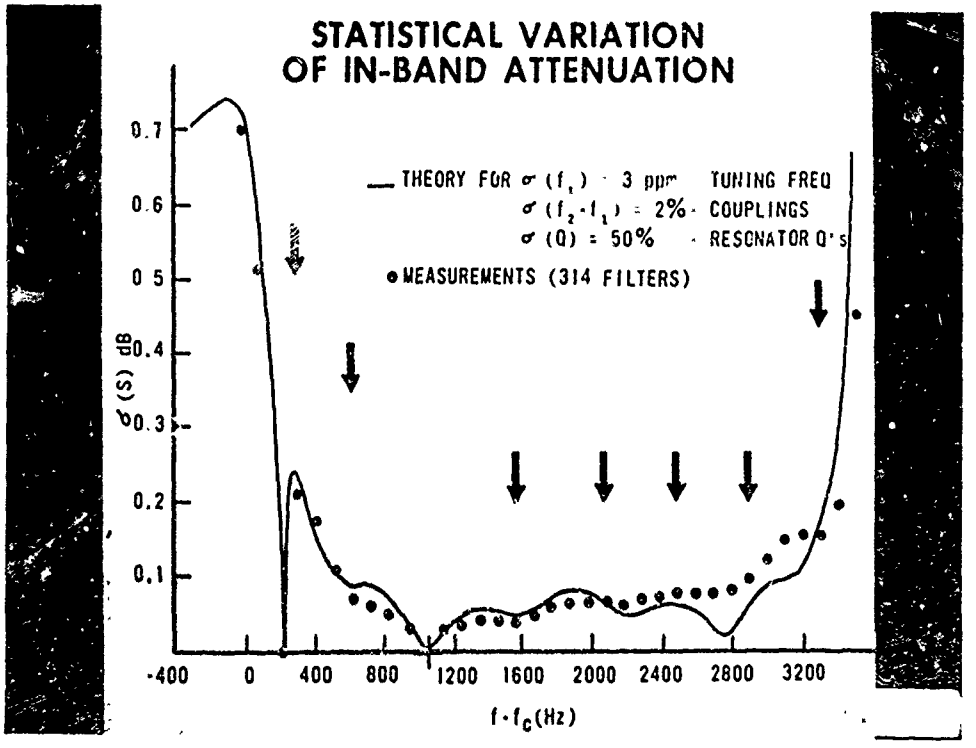
OPERATION	THROUGHPUT PER HOUR	RANGE	TOLERANCE
COUPLING ADJUST	50	10% DECREASE	± 0.4% (±8Hz)
FREQUENCY ADJUST	50	500Hz INCREASE	± 12Hz (± 1.5 PPM)
COUPLING AND FREQUENCY ADJUST	30	—	—



BLOCK DIAGRAM COMPUTER CONTROLLED TRANSMISSION MEASURING SET




A6 CHANNEL FILTER REQUIREMENTS
• AVERAGE OF 314 FILTERS
THEORETICAL 8POLE CHEBYSHEV FILTER
WITH .1dB RIPPLE, 3260Hz BANDWIDTH
AND Q=250,000
RELATIVE LOSS - dB
FREQUENCY FROM CARRIER KHz


STATISTICAL VARIATION
OF IN-BAND ATTENUATION
THEORY FOR σ(f1) = 3 ppm TUNING FREQ
σ(f2-f1) = 2% COUPLINGS
σ(Q) = 50% RESONATOR Q's
• MEASUREMENTS (314 FILTERS)
σ(S) dB
f - fc(Hz)

COMPOSITE FILTER STRUCTURES INCORPORATING MONOLITHIC CRYSTAL FILTERS AND LC NETWORKS

H. A. Simpson, E. D. Finch, Jr., and R. K. Weeman
Bell Telephone Laboratories, Incorporated
North Andover, Massachusetts

Summary

The monolithic crystal filter may be considered from either of two points of view: 1) as a complete, self-contained filter entity; or 2) as a two-port device which may be incorporated as a building block into some larger frequency selective network. The present paper considers the monolithic crystal filter from the latter point of view, with application to problems in bandpass and band-elimination filter design. For the bandpass case, this paper is principally concerned with the problem of realizing Chebyshev passband behavior from a cascade of two or more symmetrical monolithic crystal filters (MCFs) coupled by LC networks. For band-elimination filters, the theory of operation of both inductor-derived and capacitor-derived monolithic crystal filter structures is presented. Examples of practical designs for both bandpass and band-elimination filters are used to illustrate design techniques based on closed-form solutions.

Building Blocks for Composite Multi-resonator Crystal Filters

The building blocks considered in this paper are illustrated in Fig. 1. They include the following:

1. Symmetrical, coupled two-resonator crystal devices (MCF_S) having the equivalent circuit shown in the figure.
2. Narrowband LC networks of two types:
   a) Impedance inverters having ±90° phase shift; and
   b) Impedance matching networks of the "L" or "Pi" configuration.
3. Two-terminal, single-element reactances (inductors and capacitors).

The monolithic crystal filter devices considered in this paper are restricted to two-resonator symmetrical devices. These MCFs represent complete filters in their own right and can be specified by their own filter parameters. In the composite structure, however, this identity is lost, and the MCF becomes merely a device for realizing a desired performance - in this instance, a narrowband crystal filter having Chebyshev behavior in the passband. As will be seen later, Butterworth parameters permit convenient description of the individual two-resonator coupled devices.

The impedance inverter networks shown in Fig. 1 are composed of positive and negative capacitors. These are realized in practice as LC coupling networks, based on a narrowband approximation in which the positive reactance of an inductor is used to realize the reactance of a negative capacitor. As a matter of interest as will be shown later, the equivalent circuit of the MCF includes an inverter having the required positive and negative capacitances provided by the acoustic coupling between resonators.

Impedance matching networks are required at the input and output ports of the composite filter to provide an impedance match to the source and load impedances. This is necessary because the source and load impedances are generally fixed at a value different from the optimum impedance level of the composite filter.

The discrete components shown hardly need explanation. They are listed here, however, as two-terminal building blocks required for realization of inductor-derived and capacitor-derived monolithic band-elimination filters.

Some of the features of the composite filter approach to narrow bandpass filter design may be summarized as follows:

1. It allows the use of symmetrical MCF designs.
2. The coupling and impedance matching networks can be used to provide additional discrimination against overtone passbands.
3. Additional degrees of freedom are available in the coupling networks for adjustment, thus allowing requirements on the MCFs to be relaxed.
4. For small volume production, high-order crystal filters may be produced quickly - as contrasted to development times required for single plate designs.

Types of Monolithic Crystal Filters

Figure 2 shows the wide variety of MCFs that have been developed for high frequency filter applications to radio and wire systems. Since 1966, over 100 designs have been developed. Of these, approximately 12 designs of the type shown at the top of the display and identified as multi-order MCFs are for application to a telephone channel bank.

The two lower levels in the display show what are commonly called second-order or two-resonator MCFs, although they have fourth-order transfer functions. Such MCFs find application, individually, as single-tone pick-off filters or as components in a composite filter structure. The frequency ranges indicated for each header size are only approximate.

Design Parameters of Two-resonator Devices

The mathematical analyses to follow employ the equivalent electrical network as shown in Fig. 3 for the MCF. The element values of the components comprising this equivalent network completely determine the mathematical model that is used to represent the coupled resonator device. These element values are given on the left-hand side of the figure in terms of the network design parameters: R_o, the impedance

level of the filter; q_1, the normalized loaded Q of the resonators; $Bw_{3\ dB}$, the 3-dB bandwidth of the two-resonator device operated as an individual filter; f_o, the center frequency of the filter; and a device parameter n, the order of overtone operation (n=1 for the fundamental mode). The right-hand side of Fig. 3 relates the network and device design parameters. The device parameters are: A, the electrode area; t, the plate thickness; ε, the material permittivity; and K_{12}, the actual coupling between resonators in terms of the frequencies of symmetric (fs) and anti-symmetric (fa) thickness shear displacements.[1]

The network representing the acoustic coupling between resonators is electrically an impedance inverter that has the property of presenting an impedance $Z = \frac{k^2}{Z_{load}}$ at the input port when the output port is terminated in Z_{load}. The inversion constant, $k^2 = \left|\frac{1}{\omega C_m}\right|^2$, is approximately constant over the narrowband passband of the filter. Furthermore, since the z parameters Z_{11} and Z_{22} of an impedance inverter are zero, mesh resonant frequencies are unaltered when meshes are coupled by such a network.

Electrical Matching Networks

Figure 4 illustrates three LC narrow-band impedance matching networks that have been used extensively to provide a narrowband impedance match between the MCF and the source and load impedances. Back-to-back connection of such sections provides electrical coupling networks for coupling two-resonator MCFs. The networks shown in low-pass, high-pass, and bandpass forms all match R_1 to R_2 at the frequency $\omega=\omega_o$. In all cases, R_2 must be greater than R_1. The three-element bandpass section requires an additional parameter $\Delta\omega$, the 3-dB bandwidth of the section, to uniquely specify all elements. This bandpass section is particularly useful with overtone mode MCFs where the increased discrimination of the section aids in the elimination of the fundamental passband. The low-pass and bandpass-type sections provide for a broadband absorption of the shunt capacitance C_o of the MCF by supplying capacitance abutting the MCF. The remainder of this paper will consider only the low-pass-type section, although similar results have been achieved with the other sections.

Single-frequency Resistive Matching at Junctions

A possible approach to the design of a composite four-resonator crystal filter is indicated in Fig. 5. In this instance, a resistive matching technique is used at selected junctions of the network. With a load impedance of R_1, each MCF is designed at an impedance level of R_o. Considering the coupling network in the center of the schematic as two low-pass matching sections operating back-to-back, the impedance level R_x becomes the only free parameter.

The band-widening and band-shifting properties of this network are illustrated in Fig. 6. In this figure, the insertion losses of the network, for specific values of the dimensionless parameter R_x/R_o are plotted. The specific examples selected show the response of two identical Butterworth MCFs coupled by a Pi-network at a center frequency of 8.448 MHz. Individually, each MCF was designed for a 3-dB bandwidth of 1400 Hz. For the case $R_x/R_o=0.5$, an equiripple passband having a 3-dB bandwidth of approximately 1600 Hz and 1-dB ripple is obtained. If the losses of the Butterworth filters were merely additive, a bandwidth of only 1000 Hz would result. This increase in bandwidth using an LC coupling section is approximately 60 percent. For R_x/R_o above and below 0.5, large loss bumps appear above and below the passband center frequency.

Computer Optimized Results[3]

The technique of impedance matching previously shown is not, in general, capable of producing sufficiently flat passband performance for critical applications where return loss specifications might be in excess of 26 dB. Departing from the previous concept of resistive impedance matching, additional degrees of freedom result if the elements of the input matching and coupling networks are allowed to assume new design values. Typical results that can be obtained by computer-based optimization are illustrated in Fig. 7. The parameters of the MCFs were fixed, and the overall network was constrained to a symmetrical structure. This results in a total of four free variable elements. In the computer optimization routine, the flatness of the passband was optimized over a selected band.

Although four return loss peaks were obtained, the return loss characteristic was not equi-minima, nor was the insertion loss (not observable in the scale shown) equi-ripple. The deviations from the desired four-resonator Chebyshev response were only slight and negligible for all practical purposes. A closed-form solution, giving exact Chevyshev response, however, was found to exist.

Modified Impedance Matching Approach

The first impedance matching approach to composite filter design, illustrated in Fig. 5, failed to result in four-resonator Chebyshev performance. Figure 8 illustrates a modified impedance matching approach that results in exact Chebyshev response.

The matching sections are designed to present an impedance αR_o to the MCFs. The MCFs, in turn, are coupled by an LC network, the narrowband equivalent of an impedance inverter, having an inversion constant of $(R_o/\alpha K)^2$. The constant, α, is defined in terms of the coefficients K_{12} and Q_1, which are the normalized coupling and loaded resonator Q specified by Dishal[2] for a four-resonator Chebyshev filter. K is related to the desired reflection coefficient of the composite network at center frequency. The reflection coefficient, ρ, is related to A_{max}, the maximum allowable insertion loss ripple.

Decomposition of a General N-resonator Filter into a Composite Structure

Figure 9 illustrates the steps that may be taken to decompose an n-resonator (n even), equal inductance bandpass filter into an equivalent structure for realization as a composite crystal filter. The elements in the top schematic may be computed simply from considerations

outlined by Dishal.[2] The desired coupled resonator form is that shown in the top figure. Within the structure, a number of capacitive-tee impedance inverters are evident. The actual value of source and load resistance, R, may be realized as a physical resistance or, preferably, as a resistance presented by one of the narrowband matching sections of Fig. 4. The low-pass and bandpass forms of these matching sections permit broadband absorption of the shunt capacitance C_o, which has therefore been neglected in Fig. 9. The first pair of coupled resonators may be realized as a symmetrical Butterworth MCF shown schematically in the lower figure. The 3-dB bandwidth of the Butterworth filter is related to the ripple bandwidth of the composite filter and to the coupling coefficient, K_{12}, of the n-resonator Chebyshev filter as shown. The second capacitive-tee impedance inverter is realized by the narrowband equivalent Pi-network. Following this realization procedure throughout the filter, all elements may be realized as symmetrical two-resonator MCFs or as narrowband equivalents of impedance inverters.

Figure 10 summarizes the number of building blocks and the number of different building block designs required for an n-resonator composite crystal filter. The "rounding" indicated in Fig. 10 infers that improper fractions resulting from these computations must be rounded to the next whole integer. The result of the application of the decomposition technique to the realization of a 42.496-MHz eight-resonator Chebyshev composite filter is shown in Fig. 11. As determined by the chart of Fig. 10, two different symmetrical MCF designs and two different coupling network designs (narrowband impedance inverter equivalents) are required. The schematic of Fig. 11 identifies the MCFs and components. The wideband insertion loss, passband ripple, and inband return loss are seen to correspond to the exact Chebyshev filter case. In this instance, the impedance matching networks required to match R_o of the filter to specific source and load impedances are not shown.

Band-elimination Filters

Two types of monolithic crystal band-elimination filters have been developed. These filters are derived directly from monolithic crystal bandpass filters by the simple addition of inductive or capacitive elements across the ungrounded electrodes of single MCFs. The filters have been designated as inductor- or capacitor-derived types, depending on the type of bridging element.

Both the inductor-derived and capacitor-derived band-elimination filters (BEFs) are shown in Fig. 12. For the inductor - derived type, input and output ports are coupled by an inductor having a reactance equal in magnitude to that of one of the acoustic coupling capacitors of the MCF at the rejection frequency. This rejection frequency is identical to the passband center frequency of the MCF when used in its bandpass mode. For the capacitor-derived filter, a phase reversal of the monolithic filter is required. This is achieved by the connection indicated in the schematic. The phase reversal achieved is clearly evident on comparison of the two impedance inverters representing the acoustic coupling in the equivalent networks. Band-elimination characteristics are then obtained when C_x is made equal to one of the acoustic coupling capacitors, C_m.

The band-eliminating properties of these filters may be studied in a variety of ways, but perhaps the easiest approach is the application of Bartlett's Bisection Theorem to derive the equivalent lattice network. For the inductor-derived case, the sequence of networks shown at the top of Fig. 13 illustrates the method and verifies the existence of zero transmission at the rejection frequency, f_o, from the properties of a balanced lattice. The broadband equivalent lattice network for the inductor-derived filter is shown in the sequence at the bottom of Fig. 13.

For practical applications, the inductor-derived form of this BEF has superior passband performance as compared to its capacitor-derived counterpart. This is clear when it is observed, as in Fig. 14, that the BEF is essentially comprised of two filters connected in parallel: 1) a low-pass filter (C2, L_x, C2), and 2) a monolithic crystal filter. The rejection properties at f_o have already been explained. In the passband, the impedances of the individual resonators comprising the MCF are extremely high, and the resulting network reduces to a simple low-pass filter (C2, L_x, C2). These elements, and the associated source and load impedances, determine the cut-off frequency of the low-pass filter which limits the highest frequency that may be passed in a particular design. Furthermore, this impedance level cannot be selected arbitrarily in practical designs because of limitations in broadband transformer capabilities. Some compromise and iteration may be necessary at this stage to achieve a satisfactory compromise between the low-pass filter design and the design of the MCF. One must also realize that vibrational modes other than the main mode exist in the MCF and produce passband irregularities. Design considerations related to spurious mode suppression confine the MCF parameters to a limited range. The design procedure outlined in Fig. 14 has been found useful. Some iteration, however, is generally required before a satisfactory overall design is obtained.

An example of the application of these design principles to the realization of a 42.88-MHz pilot-blocking filter is shown in Fig. 15. Autotransformers are incorporated for broadband impedance matching to the 75-ohm source and load. Two inductor-derived BEF sections are required to meet the stopband requirements, and a four-section low-pass filter is chosen to permit compensation for the distributed capacitance of the transformers and for the static capacitances of the MCFs. To achieve the passband indicated as flat to within ±0.05 dB from 1 to 75 MHz, an additional amplitude equalizer is necessary. A fundamental mode MCF was required to meet passband specifications.

Conclusion

In this paper, the authors have treated the MCF as a device specifically designed as a component or building block of some larger frequence selective network. Applications to both bandpass and band-elimination filters have been cited. The theory of operation has been discussed and pertinent design techniques have been presented. Certain advantages in the

building block approach have been pointed out.

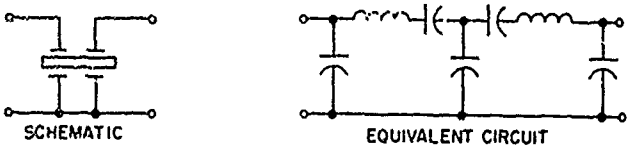
References

1. Sykes, R. A., Smith, W. L., and Spencer, W. J., "Monolithic Crystal Filters," 1967, IEEE International Convention Record, pp. 78-93.

2. Dishal, M., "Two New Equations for the Design of Filters," Electrical Communication, Vol. 30, December 1952, pp. 324-337.

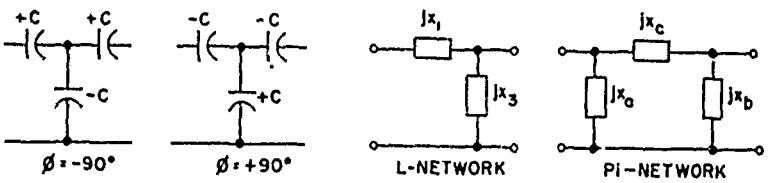
3. Garrison, J. L., Georgiades, A. N., and Simpson, H. A., "The Application of Monolithic Crystal Filters to Frequency Selective Networks," Digest of Technical Papers, 1970 International Symposium on Circuit Theory, 1970, pp. 177-178.

A. SYMMETRICAL COUPLED 2-RESONATOR DEVICES (4-TERMINAL STRUCTURE)

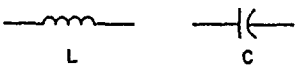


B. L-C NETWORKS (4-TERMINAL STRUCTURES)

IMPEDANCE INVERTERS IMPEDANCE MATCHING



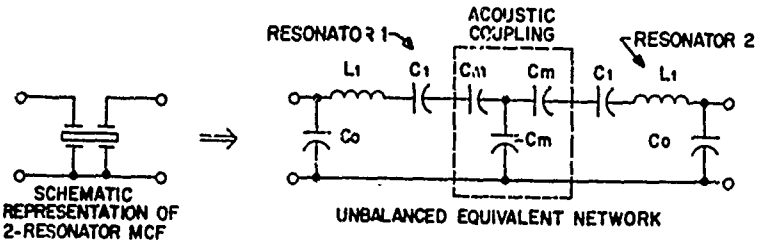
C. DISCRETE COMPONENTS (2-TERMINAL STRUCTURES)



BUILDING BLOCKS FOR COMPOSITE MULTI-RESONATOR CRYSTAL FILTERS
FIGURE 1

FIGURE 2

Family of Monolithic Crystal Filters



NETWORK DESIGN PARAMETERS

$$L_1 = \frac{R_{oq}}{2\pi B_{w3dB}} \text{ (H)}$$

$$C_1 = \frac{1}{\omega_0^2 L_1} \text{ (F)}$$

$$C_m = \frac{C_1 f_o}{k_{12} B_{w3dB}} \text{ (F)}$$

$$C_o = r n^2 C_1 \text{ (F)}$$

$$r = 250$$

DEVICE DESIGN PARAMETERS

$$L_1 = \frac{1.8n^3}{A(Cm^2) f_o^3 (MHz)} \text{ (H)}$$

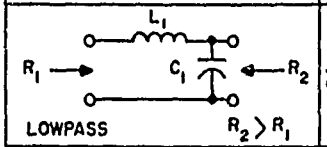
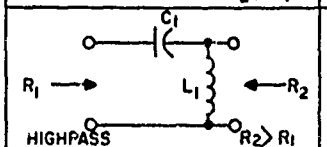
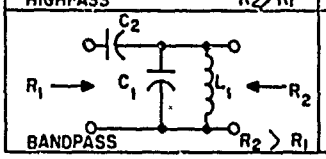
$$C_1 = \frac{1}{4\pi^2 f_o^2 L_1} \text{ (F)}$$

$$K_{12} = \frac{f_a - f_s}{\sqrt{f_a f_s}} ; C_m = \frac{C_1}{K_{12}} \text{ (F)}$$

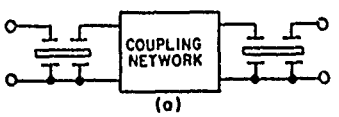
$$C_o = \frac{\epsilon A}{t} \text{ (F)}$$

$$t = \frac{0.1665n}{f_o (MHz)} \text{ (cm)}$$

EQUIVALENT NETWORK AND DESIGN PARAMETERS: 2-RESONATOR MCF
FIGURE 3

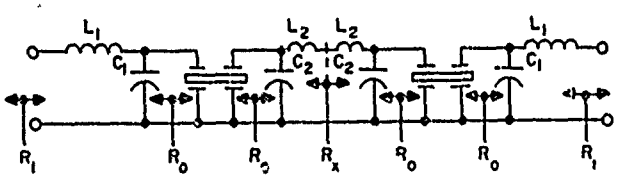
NETWORK	ELEMENT VALUES		
	L_1	C_1	C_2
 LOWPASS $R_2 > R_1$	$\frac{1}{\omega_0}\sqrt{R_1(R_2-R_1)}$	$\frac{1}{\omega_0 R_2}\sqrt{\frac{R_2}{R_1}-1}$	—
 HIGHPASS $R_2 > R_1$	$\frac{R_2}{\omega_0}\sqrt{\frac{R_1}{R_2-R_1}}$	$\frac{1}{\omega_0\sqrt{R_1(R_2-R_1)}}$	—
 BANDPASS $R_2 > R_1$	$\frac{R_2 \Delta\omega}{2\omega_0^2}$	$\frac{1}{\omega_0^2 L_1} - C_2\left(1-\frac{R_1}{R_2}\right)$	$\frac{1}{\omega_0\sqrt{R_1(R_2-R_1)}}$

NARROW BAND IMPEDANCE MATCHING NETWORKS
FIGURE 4



(a)

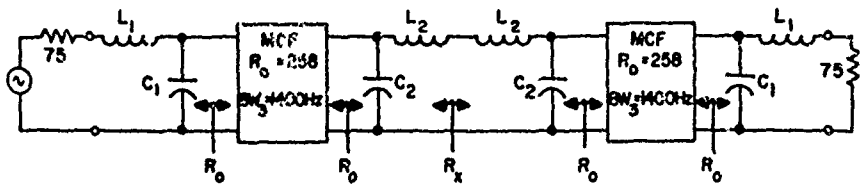
2 IDENTICAL 2-RESONATOR MCF'S & COUPLING NETWORK

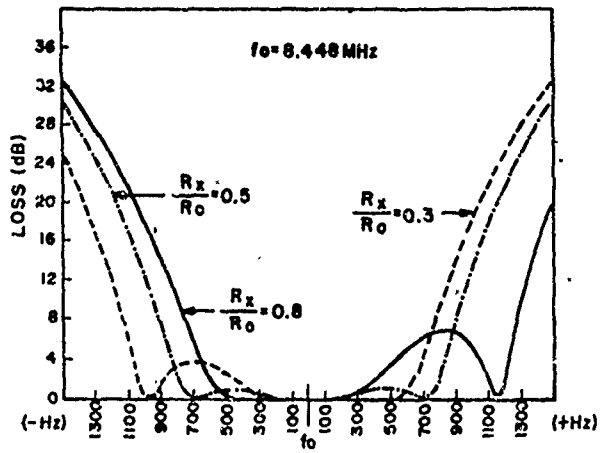


(b)

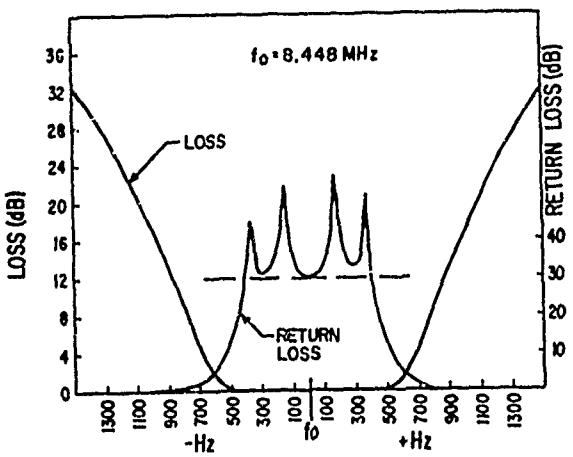
TWO IDENTICAL 2-RESONATOR MCF'S WITH L.P. IMPEDANCE MATCHING & COUPLING NETWORKS (SINGLE-FREQUENCY RESISTIVE MATCHING AT JUNCTIONS)

FOUR-RESONATOR MONOLITHIC CRYSTAL FILTERS
FIGURE 5

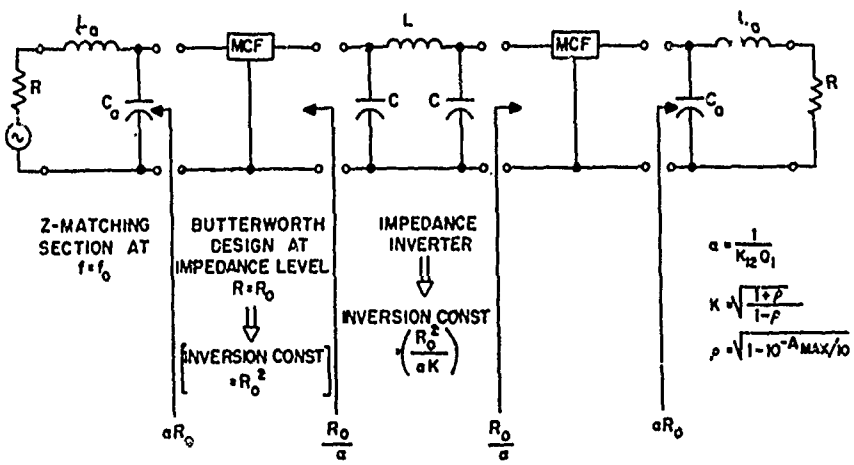




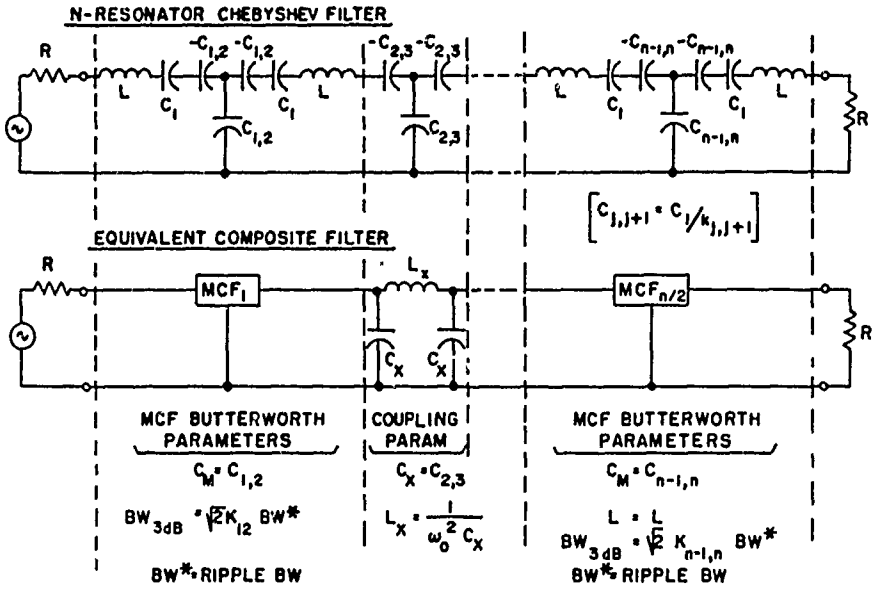
BAND-WIDENING & BAND-SHIFTING EFFECTS OF COMPOSITE NETWORK
FIGURE 6



COMPUTER OPTIMIZED RESPONSE OF COMPOSITE 4-RESONATOR FILTER
FIGURE 7



MODIFIED IMPEDANCE MATCHING REALIZATION
OF EXACT 4-RESONATOR CHEBYSHEV RESPONSE
FIGURE 8

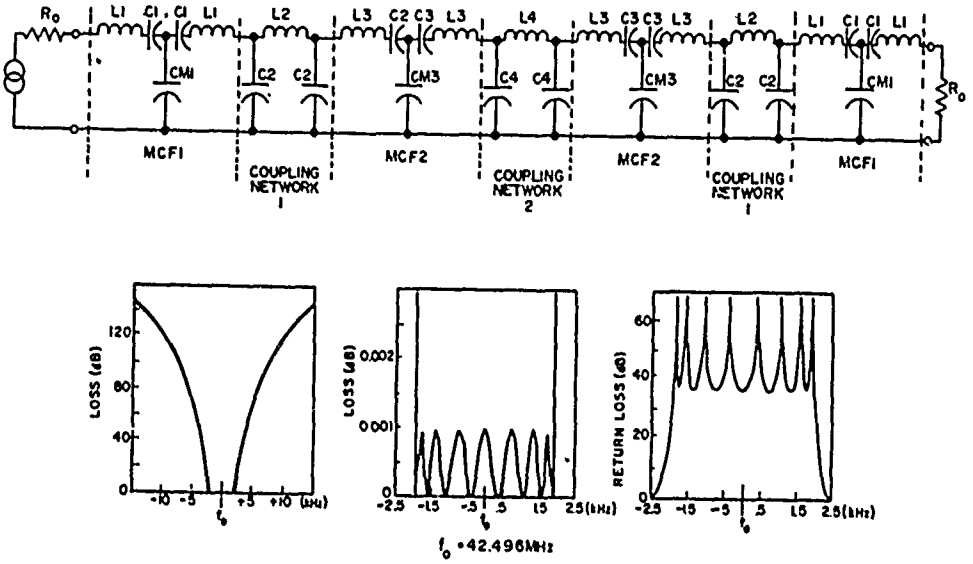


DECOMPOSITION OF n-RESONATOR FILTER TO COMPOSITE STRUCTURE
FIGURE 9

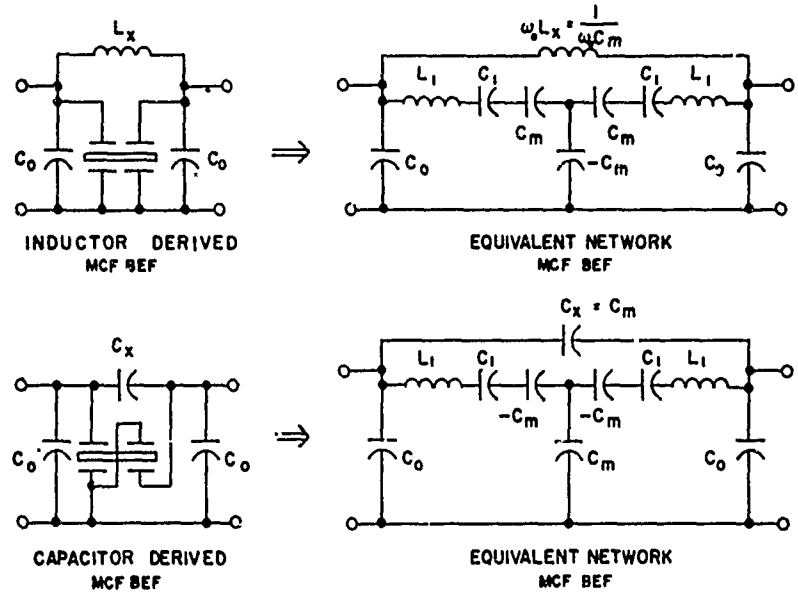
NO. OF RESONATORS	NO. OF MCF'S	NO. OF COUPLING NETS	NO. OF MCF DESIGNS	NO. OF COUPLING NET. DESIGNS
4	2	1	1	1
6	3	2	2	1
8	4	3	2	2
10	5	4	3	2
(n)	(n/2)	(n/2-1)	(n/4)*	$\left(\frac{n/2-1}{2}\right)$*

*ROUNDED

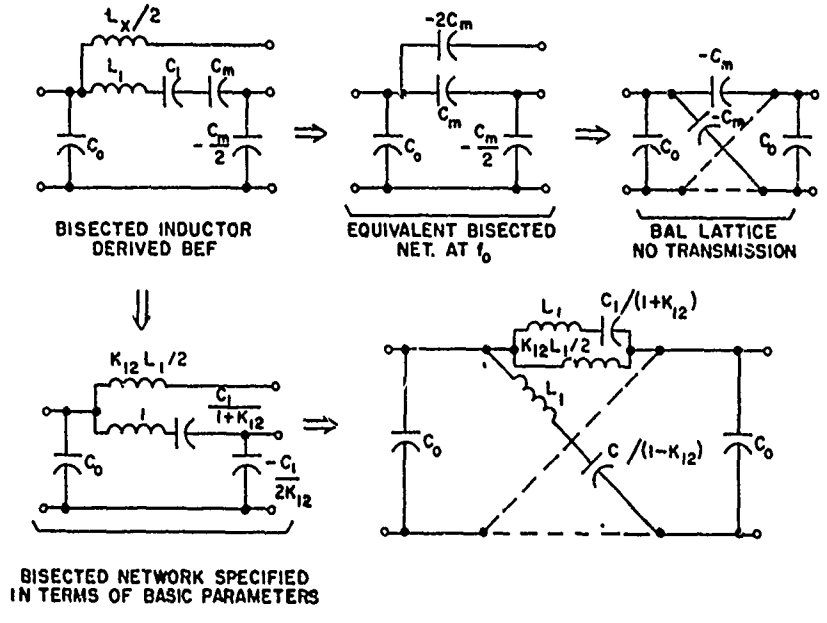
BUILDING BLOCKS REQUIRED FOR N-RESONATOR COMPOSITE FILTERS
FIGURE 10



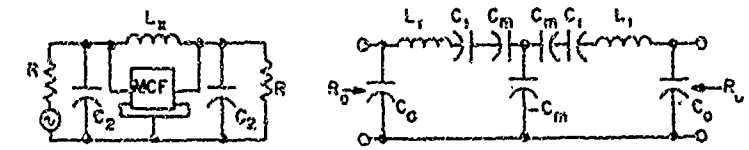
COMPOSITE EIGHT-RESONATOR CRYSTAL FILTER
FIGURE 11



EQUIVALENT NETWORKS: INDUCTOR & CAPACITOR-DERIVED MCF BEF'S
FIGURE 12



LATTICE EQUIVALENTS OF INDUCTOR-DERIVED MONOLITHIC B.E.F.
FIGURE 13

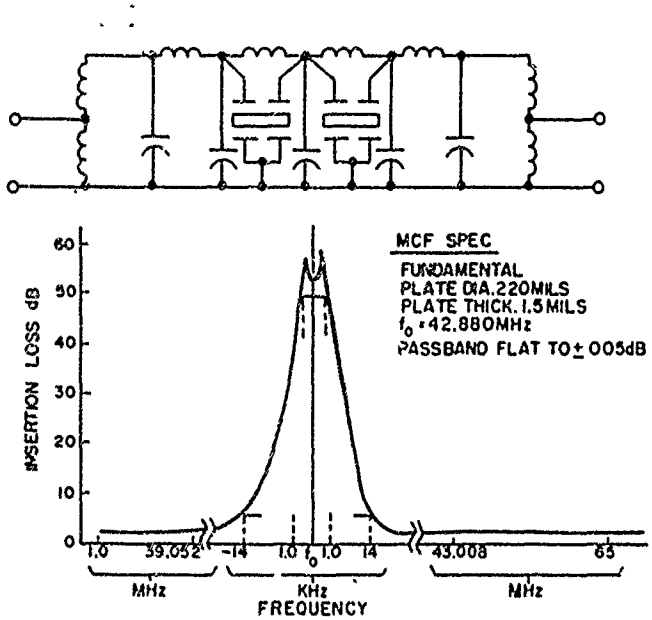


INDUCTOR-DERIVED MCF (BEF) EQUIVALENT NETWORK MCF (BPF)

DESIGN PROCEDURE

1. SELECT SUITABLE IMPEDANCE LEVEL R
2. SELECT SUITABLE f_c FOR L.P. FILTER
3. COMPUTE $L_x = 2R/\omega_c$
4. COMPUTE $C_m = 1/(\omega_0^2 L_x)$ ω_0 = SUPPRESSION PEAK FREQUENCY
5. COMPUTE $R_0 = 1/(\omega_0 k_{12} q_1 C_m)$ $k_{12} = 1/\sqrt{2}$; $q_1 = \sqrt{2}$ FOR BUTTERWORTH
6. COMPUTE $L_1 = R_0 q_1 / (2\pi \Delta W_{3dB})$
7. COMPUTE $C_1 = 1/(\omega_0^2 L_1)$
8. COMPUTE $C_3 = 1/(\omega_c R)$; $C_3 > C_0$; $C_0 = n^2 r C_1$, $C_3 = C_0 + C_2$

DESIGN PROCEDURES- MCF BAND-ELIMINATION FILTERS
FIGURE 14



INSERTION LOSS OF A 42.880 MHz PILOT BLOCKING FILTER
FIGURE 15

CESIUM ATOMIC BEAM FREQUENCY STANDARDS: A SURVEY OF LABORATORY STANDARDS DEVELOPMENT FROM 1949-1971 *

R. E. Beehler
Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80302

Summary

A general review is presented of progress achieved by the more active standards laboratories in developing laboratory cesium beam standards during the 1949-1971 period. For each significant time period during this 22-year interval an attempt is made to point out the basic approaches employed, principal characteristics of some of the more important devices developed, main problems encountered, some solutions attempted in meeting these problems, and the main accomplishments. For background purposes a brief discussion of the differences in approach between commercial and laboratory cesium standards and a general review of cesium beam operating principles are also included.

Key Words: Atomic frequency standards, Cesium beam frequency standards, Laboratory cesium standards.

Comparison of Laboratory and Commercial Cesium Standards Approaches

The purpose of this paper is to review progress achieved during the past 22 years in the development of cesium atomic beam frequency standards. The emphasis will be primarily on developments in the national standards laboratories, as contrasted with those achieved by commercial firms. For more information regarding some of the latest developments in commercial cesium standards, the reader is referred to the paper in this volume by R. Hyatt, et al. [1]

Figure 1 illustrates some of the differences between the approaches followed by commercial firms and the national standards laboratories in designing, building, and using cesium beam frequency standards. Clearly, developers of both commercial and laboratory standards strive for high accuracy and high stability. these characteristics, obviously needed by national standards labs, also make good commercial sense. However, the commercial developer must operate within constraints imposed by the necessity to produce a standard of reasonable size, weight, and reliability which will provide good performance over the rather wide range of environmental conditions found in typical field applications. Because most time and frequency applications, such as the proposed time and frequency collision avoidance system, for example, require stable sources rather than accurate ones, commercial compromises are more often made with respect to accuracy than stability. In the case of laboratory standards, on the other hand, accuracy is of paramount importance and such factors as size, weight, electrical power consumption, and so forth, are not permitted to compromise the accuracy achieved. Most laboratory standards are designed not only to provide a high accuracy capability but also to allow for thoroughly evaluating and documenting this accuracy as often as is required.

A further difference between the commercial standards laboratory approaches shows up when one examines the respective design features for achieving good frequency stability performance. One way to express this quality of stability is the widely used concept of figure of merit, which for cesium beam tubes is proportional to the signal-to-noise ratio of the cesium beam tube output signal divided by the resonance linewidth. In commercial developments a high figure of merit--and hence a good stability performance--is achieved primarily by designing for a very high beam signal-to-noise ratio. In the laboratory standards case--at least until rather recently--the designer usually tried to achieve a high figure of merit by means of a very narrow resonance linewidth--a technique that resulted in some very long beam tubes as we shall see later.

These differences in approach between commercial and laboratory standards appear to be decreasing, however, in recent years. Careful design and attention to detail are resulting in commercial tubes with stability performance comparable to that of recent laboratory standards. At the same time, factors which affect accuracy in commercial cesium standards are receiving closer scrutiny. Conversely, present-day designs of laboratory beam tubes are making increasing use of design techniques and technologies--for example, gettering materials to improve the vacuum--developed or used first by the commercial designers. One laboratory device, now being designed in Canada, is actually intended for essentially continuous use in a time scale application--much like its commercial counterparts. [2] At NBS, we have tried to obtain the best of both the commercial and standards laboratory worlds by contracting with Hewlett-Packard to jointly design and construct a refined beam tube, known as NBS-X4. This increasing cross-fertilization of ideas between commercial and standards laboratories is most encouraging; it seems quite clear already that accelerated improvement of both commercial and laboratory standards is taking place as one direct result.

In the remaining part of this paper the discussion will mainly concern laboratory standards developed during the past 20 years or so at five of the more active laboratories in Europe and North America as shown in figure 2. This is not to imply, however, that other labs

* Contribution of the National Bureau of Standards, not subject to copyright.

have not in the past or are not now also contributing to the general progress being made in this field.

Review of Cesium Beam Operating Principles

To provide a brief refresher on how a cesium beam standard, commercial or laboratory, operates, figure 3 shows one common form of the basic technique. Cesium atoms effuse from a source, shown on the left; travel down an evacuated tube at thermal velocities through a series of magnetic fields; and depending on their exact path through the tube they either strike the detector on the right side of the slide or are lost from further consideration. Atoms following either of the trajectories shown have their magnetic moments aligned by the first deflecting magnet, called the A magnet, in such a direction that forces are exerted forcing the atoms back toward the axis of the tube. As the atoms next pass through the excitation cavity region, a resonance condition exists if the microwave field in the cavity has precisely the same frequency as that corresponding to the separation of two possible energy states of the cesium atom. If the frequencies match, and if certain other conditions regarding the microwave power level and directions of the fields are satisfied, the atom will change from one of its energy levels to another, effectively causing a 180° change in the orientation of its magnetic moment as compared to its previous direction in the A magnet field. When the atom now enters the B deflecting magnet, which is identical to A, it is deflected in the opposite direction as in the A field, since its magnetic moment has been reversed in the process of making a transition between energy states. As the trajectories indicate, these atoms then strike the detector where they are ionized, producing an electrical current proportional to the number of atoms per second hitting the detector. As we have seen, this number is also proportional to the number of atoms per second making a transition and is thus a measure of whether the oscillator frequency providing the microwave field in the cavity matches the resonance frequency characteristic of the cesium atom. The detected signal, if combined with suitable servo electronics, can then be used to automatically and continuously correct the slave oscillator frequency to a constant value, related in a known way to the cesium frequency. A small, uniform DC magnetic field terms the C field, is also provided in the center region between the A and B deflecting magnets in order to keep the magnetic moments of the atoms oriented properly and to make it possible to utilize only one particular microwave energy state transition from among the group of 21 that are theoretically possible with the cesium atom.

Influences on Stability and Accuracy Performance

As background for the upcoming discussion of stability and accuracy achievements and limitations, it may be worth noting some of the aspects of this magnetic resonance technique that influence these performance measures. The stability, as we have already seen, depends on the resonance linewidth and the beam signal-to-noise ratio. It is characteristic of the technique that the resonance linewidth becomes narrower as the time of flight of the atoms between the two arms of the U-shaped excitation cavity becomes longer. The signal-to-noise ratio of the beam depends mainly on how many atoms per second make transitions and are processed by the detector, and this, of course, depends in turn on the detailed design of the beam optics system. The signal-to-noise ratio is also degraded by any background level of cesium seen by the detector, since such atoms contribute shot noise to the detection process without adding signal component. Finally excessive noise levels in the various electronic systems involved can also degrade the output stability.

The accuracy performance of a cesium standard, on the other hand, is generally degraded by any effect which causes the actual working cesium frequency as provided by the atomic beam machine to be shifted away from the ideal frequency associated with a completely unperturbed, isolated cesium atom. It is important to realize, however, that often it is only the uncertainties in the exact amount of these frequency shifts that contribute to inaccuracy. In principle, a frequency shift whose magnitude and direction are perfectly known can be compensated for by applying a correction to all measurements.

One source of error in this magnetic resonance technique involves frequency shifts produced by unwanted and unknown phase differences between the microwave fields in the two ends of the excitation cavity. A second possible problem area is uncertainties in our knowledge of the C field--that is, its magnitude, direction, and uniformity over the length and cross-section of the atomic beam. Further errors may result from uncertainties in first and second order Doppler shifts which involve knowledge of the velocity distribution of the atoms in the beam. The electronics systems can contribute a variety of errors from sources such as second harmonic distortion of the modulation signal, an asymmetrical RF spectrum used to excite the cesium resonance, and miscellaneous effects related to the servo system electronics. As implied earlier, the detailed evaluation of such accuracy-limiting effects as these for each particular cesium standard built is one of the principal activities in standards labs working with state-of-the-art devices.

Early History of Laboratory Cesium Standards

This magnetic resonance technique being reviewed is neither very new nor restricted in usefulness to atomic frequency standards. In fact, the method, without some of today's refinements, was actually perfected by Dr. Isidor Rabi at Columbia University back in the 1930's as a tool for studying atomic and molecular physics.[3] It apparently wasn't until 1945 that Dr. Rabi first suggested that this technique should be useful in building an atomic frequency standard.[4] About four years later, NBS--with the help of Professor Kusch, also from Columbia--started work on the first direct application of Rabi's technique to frequency standards. In 1952 the standard shown in figure 4, NBS-I, was successfully operated in the sense of producing the $4,0 \longrightarrow 3,0$ resonance curve of cesium-133, using only a single excitation region[5] rather than the separated field method developed by Professor Ramsey two or three years earlier.[6] Within a few months NBS-I had been modified to use the new Ramsey technique and a narrow cesium resonance only 300 Hz wide was observed.[7] At this point, Dr. Harold Lyons of NBS predicted that accuracies of 1×10^{-10} appeared possible with this type of device. NBS-I was not used during this very early period as a routine frequency standard for regular calibration work, but rather was considered a research system for making further studies and

improvements of this new type of standard. By 1955 other cesium standards were also coming on the scene. At MIT, Zacharias, Yates, and Haun were the first to develop servo systems for electronically locking a crystal oscillator to the cesium resonance.[8] Their work later led to the first commercial cesium beam standard--National's Atomichron. At NPL, Dr. Essen and his colleagues became the first group to build and place into routine operation a cesium standard for periodic calibration of secondary oscillators.[9] Figure 5 is a picture of Essen's original standard, NPL-1. It was this machine which generated the cesium-referenced measurements during the 1955-1958 period in cooperation with the U. S. Naval Observatory that led to the famous 770 number referring the cesium frequency to Ephemeris time.[10] That number has, of course, now staked out its claim to further immortality since the redefinition of the second in 1967.*

Development of Laboratory Cesium Standards: 1955-1971

In order to discuss the development of laboratory cesium standards after 1955 the next 16-year period will be divided into three separate time periods, each one of which covers a significant phase of development in getting to where we are now. The first period covers the original group of standards from four of the major labs, including the early NBS and NPL work already mentioned. The second period covers a group of second-generation standards from the same labs. Finally, the third significant period starts about 1967 when some newer approaches began to be explored and continues up to the present--and perhaps even a year or so beyond.

In the remaining portion of this paper each of these three periods will be examined in more detail--certainly not on a standard-by-standard basis, but rather from the point of view of looking for the general approaches, accomplishments, and problems coming out of each period spanning several years.

Early Development Period: 1955-1958

Beginning with the 1955-1958 period, the general approach, in simplest terms, was to build an atomic frequency standard, using the magnetic resonance technique, that would work! Little thought was given to refinement; the important thing was to develop an operating model so that studies could be made of its basic advantages and limitations. In other words, it was a necessary period in which laboratories gained valuable experience in the application of this new technique.

Since all of the labs were starting from nearly the same base level of experience, it's perhaps not too surprising that one can fairly easily identify some features which were common to all or at least most of these earliest devices. First, they all used the Ramsey technique to obtain narrower linewidths. Probably due to a general lack of experience, the beam optics systems were kept quite simple and very closely related to what had been used successfully in earlier work with the magnetic resonance technique. Since relatively small beams were used with some fairly inefficient beam optics, the resulting beam signal-to-noise ratio was rather low. Phase shift errors, if they were seriously evaluated at all, were measured by physically reversing the microwave cavity and observing a resulting change in direction of any phase shift error present. Resonance linewidths were typically a few hundred hertz wide. With the exception of the standard at LSRH these early standards did not include servo electronics, but rather required an operator to manually plot out resonance curve after resonance curve in order to calibrate an oscillator.

In spite of many problems there were some very significant accomplishments. For one thing, independent standards were constructed in at least four different laboratories which operated successfully and proved the value of the basic technique for standards applications.[9, 11-13] The accuracies achieved within a range of a few parts in 10^{11} to one part in 10^{10} may sound a little crude to laboratory standards people today, but in 1958 these results were little short of remarkable. And, as already noted, the cesium frequency was carefully related to the best known astronomical time unit.[10]

An important outgrowth of this early period was well-documented information regarding the main limitations which had to be overcome for improved performance. Some of these are listed in figure 6. The first three primarily limited the short and medium-term stability of the standards and indirectly also their accuracy, while the latter two problems affected more the very long-term stability and the accuracy evaluation of the devices. As one further example of a product of this early development period, CsI, developed at NRC in Canada, is shown in figure 7.

Second-Generation Laboratory Standards: 1959-1966

The second major time period identified--from 1959 to 1966--might be termed the "age of refinement." The general approach to building better laboratory cesium frequency standards was to refine the techniques and hardware that worked in the earliest models just discussed. Better frequency references--in the form of more stable crystal oscillators and the early atomic standards--became available and greatly aided in the evaluation of the second-generation standards. For example, at NBS, NBS-I was used as a stable reference for evaluative measurements on NBS-II, and NBS-II later proved most helpful in the evaluation of NBS-III. As understanding of the basic strengths and weaknesses of the various devices increased during this 1959-1966 period, so did the appreciation of the need for more refined evaluation techniques and measurements in order to fully document the improved accuracy and stability being observed.[14]

In terms of stability improvements during this period the trend was mainly toward designing for narrower resonance linewidths, thereby improving the figure of merit and thus the stability performance. The narrower linewidths were achieved by increasing the time for atoms to interact with the excitation field by building standards of greater length.[15-18] This trend toward longer lengths finally stopped at about 6 meters overall length and 3.7 meters interaction length for NBS-III, although a proposal was actually submitted to NBS at one

* In 1967 the 13th General Conference on Weights and Measures defined the "second" as: "the duration of 9192631770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom."

point for a vertical standard 17 meters long! Some efforts were also made during this period to increase the stability via improved beam signal-to-noise ratio--mainly, by employing larger beams with larger dipole deflecting magnets. The higher-quality crystal oscillators already mentioned helped to provide more stable excitation sources, while the more widespread use of electronic servo systems for controlling the excitation frequency finally eliminated the need for tedious and inefficient manual measurement techniques.

Significant improvements were made in these second-generation standards in reducing error sources identified from experience with the earlier models.[14-18] The frequency errors caused by cavity phase shifts were attacked on at least three different fronts. First, the basic cavity construction techniques were refined so that phase differences between the ends of the cavity were not as likely to occur. Careful measurements of electrical asymmetries during construction proved most helpful. Second, more laboratories began to make better evaluations of cavity phase shifts by physical reversal of the cavity. In one case at least, NBS-III, the cavity was left undisturbed but the beam was reversed by physically interchanging the oven and detector with some improvement in the reproducibility of the phase shift data. Third, a new technique was tried at MIT which employed square-wave phase modulation instead of the more common sinewave frequency modulation.[19] This technique makes use of the transient response of the atomic beam to provide an electrical signal related to any existing cavity phase difference, but so far none of the standards labs have adopted the method.

C-field errors were reduced somewhat by providing better magnetic shielding through additional separated layers and materials with superior magnetic properties. The uniformity of the C-field was improved in several labs by adopting a 4-wire field-producing structure.

Frequency shifts caused by RF spectrum asymmetries received much closer attention during this period. Various systems and techniques were developed for looking at spectrum problems, such as analysis of the beat note between two X-band excitation systems, use of an ammonia maser in a narrow band spectrum analyzer system, and measurements of the microwave power dependence of the cesium frequency, which turns out to be sensitive to RF asymmetries. Theoretical studies, particularly regarding the power dependence, have been helpful in pointing the way to better evaluation techniques for spectrum effects.[20] The sensitivity of the cesium frequency to RF spectrum effects added impetus to the development of lower-noise electronics components during this period and to a generally increased activity in designing and building multiplier chains, modulation systems, and servo components that were compatible with the increased performance beam tubes appearing on the scene. The possible accuracy-limiting errors caused by the presence of second harmonic distortion of the modulating signal are a specific example of an electronic problem that received much attention. In addition to the successful, straightforward reduction of second harmonic distortion levels through better circuit designs and construction procedures by most of the labs, a more novel approach was taken at NRC by developing a new, square-wave frequency modulation system that employed 10-second-interval switching between two appropriate microwave frequencies to successfully reduce systematic errors from the electronics.[21]

Summarizing the main accomplishments during this 1959-1966 period, a number of second-generation laboratory standards were built, most of which are still used today--though perhaps with added improvements. The accuracy performance was improved by a factor of ten to levels varying from 3×10^{-12} (1 σ) to 1×10^{-11} (1 σ). Stability improvements, largely due to the narrow 40-100 Hz linewidths achieved, resulted in performance of 1×10^{-11} (1 σ) for 1-second averaging times improving as $\tau^{-\frac{1}{2}}$ to a level of 1×10^{-13} or better. Typical tabulations of bias uncertainties showed that in spite of significant progress phase shift errors, electronics effects, and C-field uncertainties continued to be the most serious accuracy limitations.

Figures 8, 9, and 10 show three of these second-generation standards: NPL-2, NBS-III, and CsIII (NRC), respectively.

More Recent Laboratory Standards: 1967-1971

The last of the three time periods covers the period from 1967 to the present and is characterized by a continuing attempt to refine existing techniques but also by a more active development of some new techniques. It might also be considered as a period of greater sophistication where previous tendencies to design longer standards to achieve greater stability have been replaced by concentrated attempts to increase the signal-to-noise ratio of the beam. For example, the present PTB standard with an excitation region only 79 cm long has produced 1-second stabilities as low as 3×10^{-12} or a figure of merit of nearly 30 as compared to about 10 for present long-beam standards.[22] An even smaller tube being developed as LSRH has shown promise of even better stability based on signal-to-noise measurements.[23] In the case of the NBS-X4[24] and NBS-5[25] designs, a large increase in signal-to-noise ratio is made possible by using a large digital computer to optimize the parameters of the beam optics systems. NBS-X4, with a total length of about 1.3 meters, should provide a figure of merit of nearly 100, corresponding to a stability of 1×10^{-12} in 1 second. NBS-5, which is a major rebuilding of NBS-III, is designed for a figure of merit of at least 500, corresponding to a 1-second stability of about 2×10^{-13}.

Other laboratories, particularly NRC,[25] PTB,[27] and LSRH,[23] have been working with beam tubes employing hexapole or combinations of hexapole and dipole deflecting magnets to increase the useful beam intensity significantly. The primary advantage in using hexapole magnets, of course, lies in their ability to focus a cylindrical beam of atoms. In addition to the direct increase in useful beam intensity that results, a further advantage is gained indirectly by being able to operate a higher cesium oven temperatures without running into collision problems in the simpler beam collimator permitted by the hexapole design.[28]

Figure 11 shows some of the possible variations on the multipole beam optics designs. For comparison, the upper scheme is just the usual flop-in technique with dipole deflecting magnets. In the second situation--the hexapole flop-in case--atoms which make transitions are focused on an annular detector. Two difficulties with

this scheme are that the relatively large surface area of the annular detector tends to increase noise levels in the detection system and the annular detector is harder to construct. The hexapole flop-out system shown next, which is that used in the present PTB standard,[27] permits a small, spot detector, but suffers somewhat from the relatively high background of atoms which strike the detector without first making a transition. These atoms contribute to the shot noise without adding useful signal. Lastly, the bottom scheme shows the hybrid system being constructed and tested now at LSRH, featuring a hexapole A magnet and a double-gap dipole B magnet.[23] In this design the previously-mentioned disadvantages of large detector surface area and difficult construction problems in one case and a large background of unflopped atoms in the second case are all eliminated. A small price is paid, however, in terms of an intensity reduction due to the dipole B magnet's inability to process some of the atoms supplied by the hexapole A magnet. Figure 12 shows a view of this rather unique double-gap dipole B magnet used at LSRH--the center structure serves both as one of the magnet poles and a beam stop to eliminate unwanted atoms.

Good progress is also being made on the reduction of systematic errors which affect the accuracy of these newer laboratory standards. The cavity phase shift problem is being approached mainly along three lines. First, more of the standards are now being built with facilities for reversing the beam direction through the undisturbed cavity as a means of precisely and accurately evaluating this error source. Second, in the case of the existing NBS-X4 and CsI (PTB) beam tubes, the single oven and detector can be physically interchanged--in the case of the PTB standard, without breaking the vacuum. Third, the modified long-beam machine at NPL,[29] the new CsV(NRC) standard now being designed in Canada,[26] a proposed new standard at PTB,[22] and the nearly-completed NBS-5[25] are all designed with a movable oven and detector at each end to operate with a beam in either direction for phase shift evaluation quickly, easily, and under more optimum conditions. Errors from this source should then be reduced to 1×10^{-13} or below. Another more novel approach is employed in the present PTB standard where the dependence of the phase shift error on the average beam velocity is used to detect phase shift error.[27] The beam velocity can be varied in this machine over a ratio of 2:1 by interchanging two different sets of focusing magnets designed to focus widely different velocities--the interchange requires only about one second to complete.

C-field errors are still being improved in the newer standards both by better shielding designs and materials and by generating more uniform fields. As an impressive example, the present CsI standard at PTB employs a longitudinal C-field rather than the usual perpendicular-to-the-beam orientation, allowing the use of a very uniform solenoid coil structure to produce fields sufficiently uniform over the transition region to reduce associated errors to less than 3×10^{-14}.[33] Figure 13 shows a picture of this machine.

Electronics improvements seem to be generally keeping pace with beam tube progress. Some impressive results have been obtained by NBS and some commercial labs in terms of building signal processing equipment and even oscillators with greatly-improved phase noise characteristics.[25, 30] Also in the area of electronics, NRC has developed an improved version of their square-wave frequency modulation system.[31] The new system uses a faster switching rate of 8 1/3 Hz, and NRC feels that the total error contributed by the complete electronics system, including multipliers, modulation system, and servo components, is now less than 1×10^{-13}. PTB has adopted the same general type of square-wave system, featuring a 3 Hz switching rate and appropriate suppression of transient effects. The PTB group has also attempted to provide more stable excitation sources by designing their system for locking the excitation oscillator to either a rubidium cell, a hydrogen maser, or a very good crystal oscillator.

Finally, the second order Doppler shift uncertainty, which becomes very important at accuracy levels of 5×10^{-13} or better, has been studied extensively at NRC[26] and NBS.[32] The major effort is directed toward developing methods of inferring the actual velocity distribution in the atomic beam from an analysis of the observed Ramsey resonance curves. Results thus far look very promising for reducing this uncertainty to less than 1×10^{-13}.

Conclusion

Figure 14 summarizes the accuracies achieved by the various national standards labs, beginning in 1956 and including some projected values for standards now in the process of construction. The plotted values are generally equivalent to 1 σ estimates and are based on published results from the various labs discussed. The exact placement of some of the points is somewhat uncertain, since it is not always possible to determine exactly when a given accuracy was first achieved. However, it is clear that accuracies of 4×10^{-13} have already been achieved[22, 33] and that improved values of near 1×10^{-13} are expected within the next year or so.[22, 25, 26]

In terms of stability performance the best reported value so far[22, 33] is $\sigma(2, \tau) = 2.8 \times 10^{-12} \tau^{-\frac{1}{2}}$. Within the next year this value should be improved by more than a factor of ten[25] to near 2×10^{-13}.

The generally high level of performance being achieved by today's laboratory cesium standards is confirmed by analyses of long-term comparison data among the various standards by means of Loran-C and portable clock trips. The peak-to-peak spread of frequencies of the lab standards appears to be about 1×10^{-12}.[34]

Acknowledgments

Because of the review nature of this paper, it has been especially necessary to rely on contributed information and pictures from a number of people at the major standards laboratories. In particular, I would like to acknowledge the valuable assistance of Dr. L. Essen of NPL, Dr. P. Kartaschoff of LSRH, Mr. B. Fischer of PTB, Dr. A. Mungall of NRC, and Mr. D. Glaze of NBS. Without their fine cooperation some of the information presented--particularly, that concerning present status and future projections--might not have been available.

References

1. R. Hyatt, D. Throne, L. S. Cutler, J. H. Holloway, and L. F. Mueller, "Performance of Newly Developed Cesium Beam Tubes and Frequency Standards," Proc. 25th Frequency Control Symposium (1971).

2. A. G. Mungall, NRC, Ottawa, Canada, Private communication.

3. I. I. Rabi, J. R. Zacharias, S. Millman, and P. Kusch, "A New Method of Measuring Nuclear Magnetic Moments," Phys. Rev., vol. 53, p. 318 (February 1938).

4. W. Hershberger and L. Norton, "Frequency Stabilization with Microwave Spectral Lines," RCA Review, vol. 9, pp. 38-49 (March 1948).

5. J. Sherwood, H. Lyons, R. McCracken, and P. Kusch, "High Frequency Lines in the HFS Spectrum of Cesium," Bull. Am. Phys. Soc., vol. 27, p. 43 (1952).

6. N. Ramsey, "A Molecular Beam Resonance Method with Separated Oscillating Fields," Phys. Rev., vol. 78, pp. 695-699 (June 1950).

7. H. Lyons, "Spectral Lines as Frequency Standards," Ann. N.Y. Acad. Sci., vol. 55, pp. 831-871 (November 1952).

8. J. R. Zacharias, J. G. Yates, and R. D. Haun, "An Atomic Frequency Standard," Proc. IRE (Abstract), vol. 43, p. 364 (March 1955).

9. L. Essen and J. Parry, "Atomic Standard of Frequency and Time Interval," Nature, vol. 176, pp. 280-282 (August 1955).

10. W. Markowitz, R. Hall, L. Essen, and J. Parry, "Frequency of Cesium in Terms of Ephemeris Time," Phys. Rev. Lett., vol. 1, pp. 105-107 (August 1958).

11. R. C. Mockler, R. E. Beehler, and C. S. Snider, "Atomic Beam Frequency Standards," IRE Trans. Instrumentation, vol. I-9, pp. 120-132 (September 1960).

12. S. Kalra, R. Bailey, and H. Daams, "Cesium Beam Standard of Frequency," Can. J. Phys., vol. 36, pp. 1442-1443 (1958).

13. P. Kartaschoff, J. Bonanomi, and I. de Prins, "Cesium Frequency Standards: Description and Results," Helv. Phys. Acta, vol. 33, pp. 969-973 (1960).

14. R. Beehler, R. Mockler, and J. Richardson, "Cesium Beam Atomic Time and Frequency Standards," Metrologia, vol. 1, pp. 114-131 (July 1965).

15. L. Essen and J. Parry, "An Improved Cesium Frequency and Time Standard," Nature, vol. 184, p. 1791 (December 1959).

16. A. Mungall, H. Daams, and R. Bailey, "The Canadian Cesium Beam Frequency Standard," Proc. 20th Frequency Control Symposium, pp. 436-447 (1966).

17. R. Beehler and D. Glaze, "The Performance and Capability of Cesium Beam Frequency Standards at the National Bureau of Standards," IEEE Trans. Instrumentation and Measurement, vol. IM-15, pp. 48-55 (March-June 1966).

18. P. Kartaschoff, "Operation and Improvement of a Cesium Beam Standard Having 4-Meter Interaction Length," IRE Trans. Instrumentation, vol. I-11, p. 224 (1962).

19. R. Bedessa, V. Bates, and C. Searle, "Frequency-Impulse Modulation as a Means of Attaining Accuracy in Cesium Atomic Clocks," IEEE Trans. Instrumentation and Measurement, vol. IM-13, pp. 175-180 (December 1964).

20. R. Harrach, "Radiation-field-dependent Frequency Shifts of Atomic Beam Resonances," J. Appl. Phys., vol. 38, pp. 1808-1819 (March 15, 1967).

21. H. Daams, "Cesium Beam Servo System Using Square Wave Frequency Modulation," Proc. 24th Frequency Control Symposium, pp. 294-300 (1970).

22. B. Fischer, PTB, Private communication.

23. P. Kartaschoff and P. Debély, "Resonateur a Cesium de Conception Nouvelle," Proc. Colloque International de Chronométrie, Série A, pp. A3-1 - A3-18 (1969).

24. D. Glaze, NBS, Private communication.

25. D. Glaze, "Improvements in Atomic Cesium Beam Frequency Standards at the National Bureau of Standards," IEEE Trans. Instrumentation and Measurements, vol. IM-19, pp. 156-160 (August 1970).

26. A. Mungall, NRC, Private communication.

27. G. Becker, B. Fischer, G. Kramer, and E. Mueller, "Neukonstruktion Eines Cäsiumstrahl-Zeitnormals an der Physikalisch-Technischen Bundesanstalt," Proc. Colloque International de Chronométrie, Série A, pp. A1-1 - A1-12 (1969).

28. J. Holloway and R. Lacey, "Performance of Cesium Beam Tubes with Multipole Optics," Proc. Colloque International de Chronométrie, Série A, paper no. A3 bis (1969).

29. L. Essen and D. Sutcliffe, "Improvement to the National Physical Laboratory Atomic Clock," Nature, vol. 223, pp. 602-603 (August 9, 1969).

30. H. Brandenberger, F. Hadorn, D. Halford, and J. Shoaf, "High Quality Quartz Crystal Oscillators: Frequency Domain and Time Domain Stability," Proc. 25th Frequency Control Symposium (1971).

31. H. Daams, "Cesium Beam Servo System Using Square Wave Frequency Modulation," Proc. 24th Frequency Control Symposium, pp. 294-300 (1970).

32. S. Jarvis, NBS, Private communication.

33. G. Becker, B. Fischer, G. Kramer, and E. Mueller, "Diskussion der Inneren Unsicherheit des Neuen Cäsiumstrahl-Zeitnormals der Physikalisch-Technischen Bundesanstalt," Proc. Colloque International de Chronométrie, Série A, pp. A2-1 - A2-9 (1969).

34. D. Allan, D. Davis, B. Blair, and H. Machlan, "Precision and Accuracy of Remote Synchronization via Portable Clocks, Loran C, and Network Television Broadcasts," Proc. 25th Frequency Control Symposium (1971).

General approaches to Cs standards

COMMERCIAL

- SIZE, WEIGHT, POWER CONSUMPTION, RELIABILITY
- GOOD PERFORMANCE OVER WIDE ENVIRONMENTAL RANGE
- ECONOMIC CONSIDERATIONS
- STABILITY EMPHASIZED OVER ACCURACY
- GOOD FIGURE-OF-MERIT VIA HIGH BEAM S/N RATIO
- CONTINUOUS OPERATION

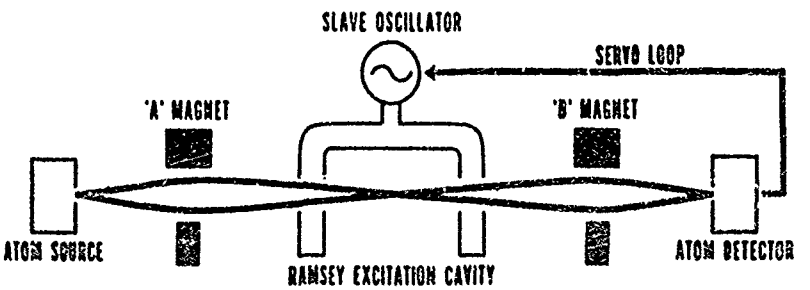
LABORATORY

- ACCURACY, EVALUATION & STABILITY
- WIDE ENVIRONMENTAL RANGE NOT IMPORTANT
- GOOD FIGURE-OF-MERIT VIA NARROW LINEWIDTH & HIGH S/N RATIO
- INTERMITTENT OPERATION

Important laboratories in development of Cs standards

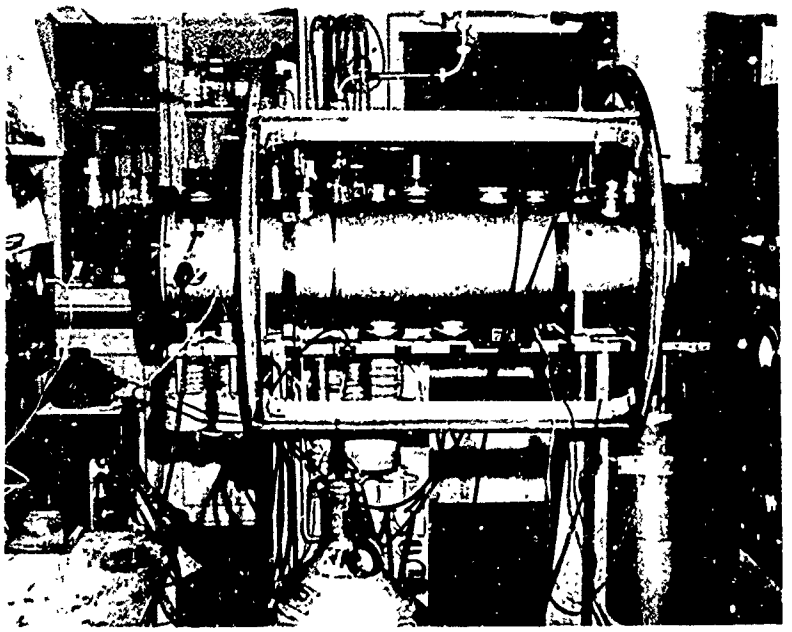
LSRH	LABORATOIRE SUISSE DE RECHERCHES HORLOGERES NEUCHATEL, SWITZERLAND
NBS	NATIONAL BUREAU OF STANDARDS BOULDER, COLORADO
NPL	NATIONAL PHYSICAL LABORATORY TEDDINGTON, ENGLAND
NRC	NATIONAL RESEARCH COUNCIL OTTAWA, CANADA
PTB	PHYSIKALISCH–TECHNISCHE BUNDESANSTALT BRAUNSCHWEIG, W. GERMANY

FLOP-IN ATOMIC BEAM

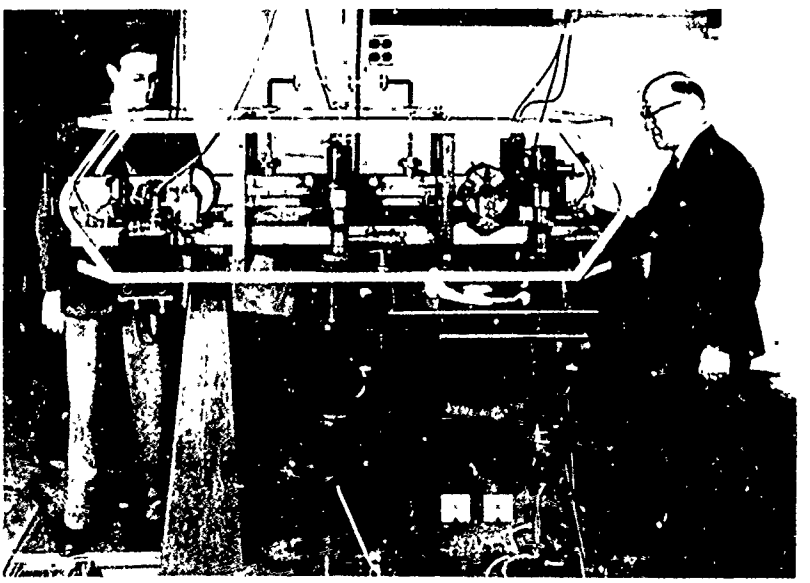


CESIUM 9 GHz

Operating principles for flop-in atomic beam technique



NBS-I cesium beam standard.



NPL [illegible] beam standard (courtesy of Dr. L. Essen, NPL
(Crown Copyright) Teddington).

Early models

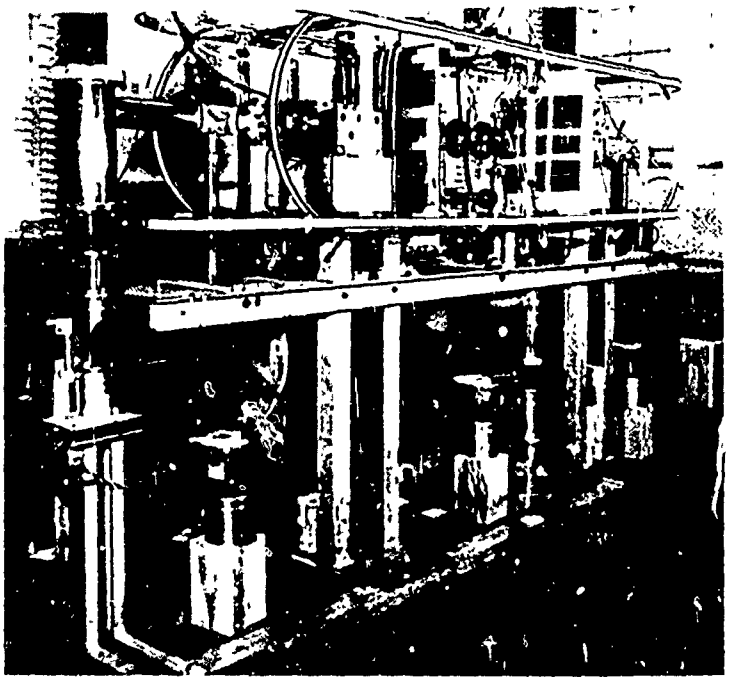
BASIC ACCOMPLISHMENTS

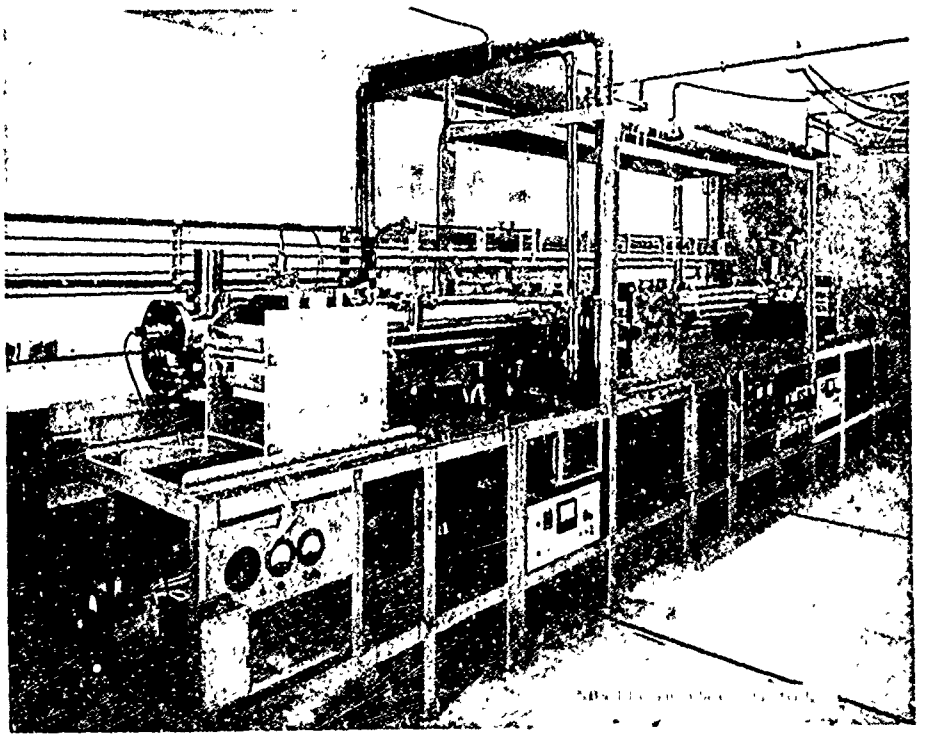
- APPLICATION OF MAGNETIC RESONANCE TECHNIQUE
- HIGH ACCURACY & STABILITY POTENTIAL
- ACCURACIES: A FEW × 10^{-11} TO 1 × 10^{-10}
- MEASUREMENT OF Cs FREQUENCY IN TERMS OF EPHEMERIS TIME

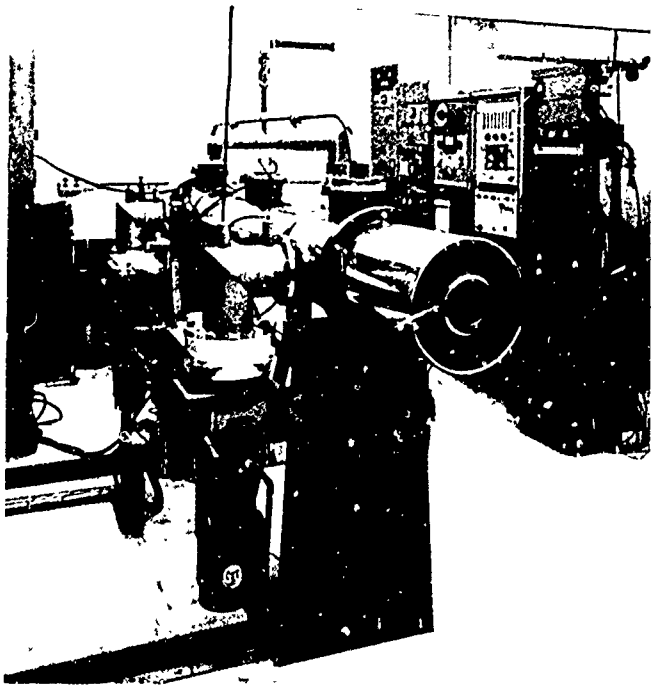
MAIN LIMITATIONS

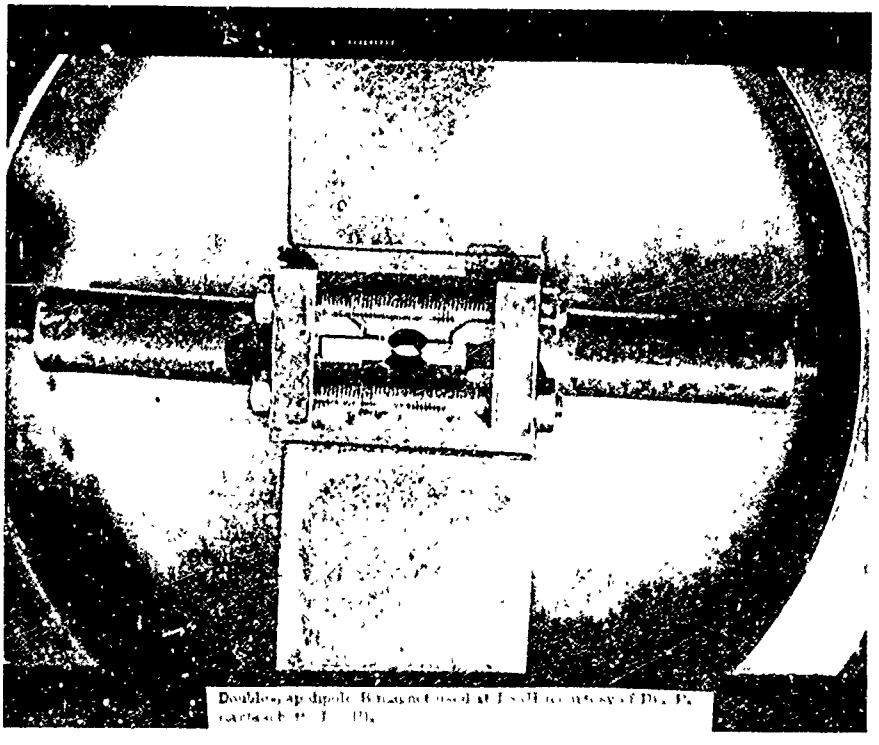
- STABILITY OF EXCITATION SOURCES
- MANUAL MEASUREMENTS
- LOW BEAM S/N RATIO
- LARGE, UNSTABLE CAVITY PHASE SHIFTS
- MAGNETIC FIELD UNCERTAINTIES & NON-UNIFORMITIES

Basic Accomplishments and limitations of early cesium beam standards



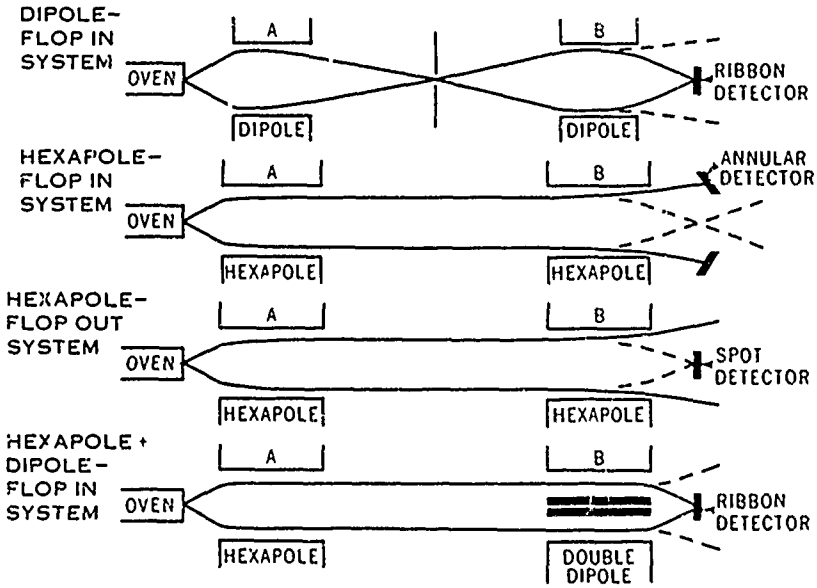




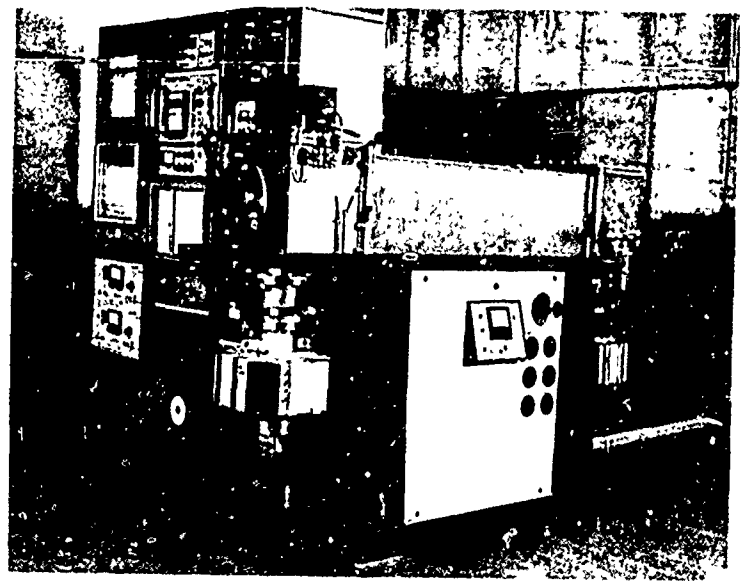


Double-spaced dipole Beacon Array used at [illegible] University of [illegible] Antarctica [illegible]

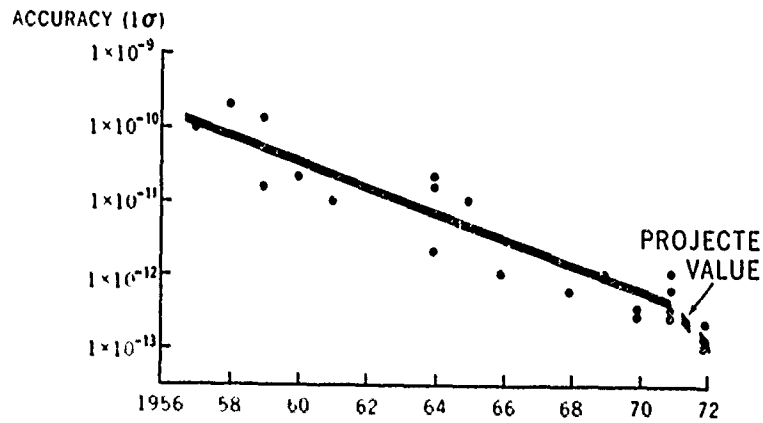
Dipole and multipole beam optics systems



ATOMS WHICH MAKE TRANSITIONS FOLLOW SOLID LINES IN ALL CASES



Accuracy trends in lab Cs standards



FREQUENCY BIASES IN A BEAM TUBE CAUSED BY RAMSEY EXCITATION PHASE DIFFERENCES

Helmut Hellwig, James A. Barnes, and D. J. Glaze

Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80302

Abstract

A phase difference between the two interaction regions of a Ramsey excitation resonance structure results in a frequency bias in the measured beam resonance. A simple mathematical model is discussed which describes the dependence of this bias on the phase difference, the microwave power level, the modulation amplitude, and the resonance linewidth. This dependence results from the interaction of the modulated microwave excitation frequency with the asymmetric shape of the slightly shifted resonance line. In a first order approximation, no dependency on the frequency modulation amplitude is expected. Near-linear dependencies on the linewidth and microwave power level which are quite pronounced even at relatively small cavity phase differences are predicted.

The theoretical results are compared with one set of experimental data on the microwave power dependence as measured in 1969 with the primary cesium beam standard NBS-III. After a correction is applied to remove the power dependence due to spectral impurity of the microwave excitation, the remaining measured power dependence agrees quantitatively with that calculated using a cavity phase difference of about 3 milliradians.

Key Words: Cavity phase shift, Cesium beam tube, Frequency accuracy, Frequency standard, Power shift, Resonance line shape.

1. Introduction

In accordance with the international definition,[1] the unit of time interval, the second, is realized with a cesium beam apparatus. Several bias corrections have to be applied to the measured resonance frequency in order to obtain the resonance frequency of the unperturbed, free cesium atom.[2] In this paper we consider only those bias corrections which are associated with a phase difference between the two interaction regions of a Ramsey excitation resonance structure.[3]

In this paper we present a simple mathematical model which allows analytical solutions to the question of frequency shifts in the presence of a cavity phase difference. The results are intended to aid in the understanding of the physical mechanisms involved. Although our mathematical model is only an approximation to the physical situation, and is tailored to a particular set of conditions, we feel that it is close enough to the real conditions which were encountered in the National Bureau of Standards Frequency Standard,[2] NBS-III, to warrant a quantitative comparison with experimental data from this apparatus. This is done in Section V of this paper.

II. Calculation of Frequency Shifts

We will proceed on the basis of the following set of assumptions and approximations: (a) No frequency deviations from resonance exceed the linewidth; (b) the microwave interrogation power does not deviate greatly from optimum power; (c) the modulation of the microwave signal is sinusoidal; (d) the modulation frequency is less than the resonance linewidth; (e) the velocity distribution in the atomic beam is Maxwellian, and (f) $l \ll L$ where l and L are the lengths of the interaction and drift regions, respectively, as defined by Ramsey.[3]

In the presence of a cavity phase difference the central peak of the Ramsey resonance pattern will be displaced from the resonance frequency of the unperturbed atom. Furthermore, the lineshape will display a certain asymmetry. We restrict ourselves to the conditions (a) and (b) above and can in first order approximate the central peak of the Ramsey pattern, $I(\nu)$, by

$$I(\nu) = A + B\phi \sin\left[\frac{\pi(\nu - \nu_\phi)}{U}\right] + C\cos\left[\frac{\pi(\nu - \nu_\phi)}{W}\right]. \quad (1)$$

In eq (1) A, B, and C are constants under given experimental conditions; W is the linewidth at half intensity; U has the characteristic of a linewidth: $U \approx W$ for monovelocity atoms, U is the same order of magnitude as W for our assumption (e) above; and ν_ϕ is a parameter related to the cavity phase difference ϕ. Using condition (e) above, it can be shown that

$$\nu_\phi - \nu_o = \frac{3\phi}{4\pi} W. \quad (2)$$

In eq (2) ν_o is the resonance frequency in the absence of a cavity phase difference. Because of the asymmetry, the frequency of the central peak ν_p differs from ν_ϕ. By setting $dI(\nu)/d\nu = 0$, we obtain

$$\nu_p = \nu_\phi + \frac{B W^2 \phi}{\pi C U}. \quad (3)$$

We now proceed to calculate the frequency to which the slave oscillator frequency servo will "tune-in." If we modulate ν in a sinusoidal fashion [condition (c) above] about some value ν_1 with the amplitude ν_m and the angular frequency ω we can write

$$\nu(t) = \nu_1 + \nu_m \sin \omega t. \quad (4)$$

We now make use of assumption (d) above and substitute eq (4) into eq (1), using eq (3) to express B. We obtain

$$I[t] = A + C\left[(\nu_p - \nu_\phi)\frac{\pi U}{W^2}\sin\left|\frac{\pi}{U}(\nu_1 - \nu_\phi + \nu_m \sin\omega t)\right| + \cos\left|\frac{\pi}{W}(\nu_1 - \nu_\phi + \nu_m \sin\omega t)\right|\right]. \quad (5)$$

We are interested in the coefficient of that term in I[t] which varies as sin ωt, because the servo forces this term to be zero.

Now

$$\sin\left|\frac{\pi}{U}(\nu_1 - \nu_\phi + \nu_m \sin\omega t)\right| = 2\cos\left|\frac{\pi}{U}(\nu_1 - \nu_\phi)\right| J_1\left|\frac{\pi\nu_m}{U}\right| \sin\omega t + \text{harmonics}$$

and

$$\cos\left|\frac{\pi}{W}(\nu_1 - \nu_\phi + \nu_m \sin\omega t)\right| = -2\sin\left|\frac{\pi}{W}(\nu_1 - \nu_\phi)\right| J_1\left|\frac{\pi\nu_m}{W}\right| \sin\omega t + \text{harmonics}.$$

Then the coefficient of sin ωt in I[t] [eq (5)] vanishes when

$$\sin\frac{\pi}{W}(\nu_1 - \nu_\phi) = (\nu_p - \nu_\phi)\frac{\pi U}{W^2}\cos\frac{\pi}{U}(\nu_1 - \nu_\phi)\frac{J_1\left(\frac{\pi\nu_m}{U}\right)}{J_1\left(\frac{\pi\nu_m}{W}\right)}. \quad (6)$$

In eq (6), ν_1 can now be interpreted as the resonance frequency of the beam tube, since ν_1 is the frequency to which the servo will "tune-in." Note that ν_1 is not equal to either ν_p or ν_ϕ.

We have assumed that the offset is always small compared to the linewidth, i.e., $(\nu_1 - \nu_\phi) \ll W$, $(\nu_1 - \nu_\phi) \ll U$, and we can rewrite eq (6)

$$\nu_1 - \nu_\phi \approx (\nu_p - \nu_\phi)\frac{U J_1\left|\frac{\pi}{U}\nu_m\right|}{W J_1\left|\frac{\pi}{W}\nu_m\right|}. \quad (7)$$

Substituting eq (2) and eq (3) into eq (7) we obtain

$$\nu_1 - \nu_o \approx \frac{3\phi W}{4\pi}\left[1 + \frac{4B}{3C}\frac{J_1\left|\frac{\pi}{U}\nu_m\right|}{J_1\left|\frac{\pi}{W}\nu_m\right|}\right]. \quad (8)$$

Equation (8) gives us the frequency bias $\Delta\nu$ which we define

$$\nu_1 - \nu_o \equiv \Delta\nu. \quad (9)$$

In order to evaluate eq (8) we have to express B/C in terms of the operating conditions. Following eq (1) we choose a linear approximation based on the numerical results of Ref. 3, and we obtain $B\phi\pi/CU \approx 2\phi/3\pi W_o$. W_o is the linewidth at optimum power P_o. Equation (8) can now be written

$$\Delta\nu \approx \frac{3\phi W}{4\pi}\left[1 + \frac{8U}{9\pi^2 W_o}\frac{J_1\left|\frac{\pi}{U}\nu_m\right|}{J_1\left|\frac{\pi}{W}\nu_m\right|}\right]. \quad (10)$$

III. Frequency Modulation Amplitude Dependence

In eq (10) any dependence on the frequency modulation amplitude is solely due to the arguments of the Bessel functions which contain ν_m. The approximate nature of our discussion requires that we restrict our discussion to modulation amplitudes not exceeding the optimum modulation, that is to

$$\nu_m \leq \nu_{mo} \equiv \frac{1}{2}W_o. \quad (11)$$

This and the fact that U and W can be expected not to differ greatly leads us in a first order approximation to

$$\Delta\nu \approx \frac{3\phi W}{4\pi}\left[1 + \frac{8W}{9\pi^2 W_o}\right]. \quad (12)$$

According to eq (12), $\Delta\nu$ has no functional dependence on ν_m. We conclude that in the limits of our approximation no frequency modulation amplitude dependent frequency bias is to be expected for $\nu_m \leq \nu_{mo}$. This result is in accordance with our experimental observations on the cesium beam frequency standards at the National Bureau of Standards.

IV. Power Dependence

We can use eq (12) to obtain the dependence of the frequency bias $\Delta\nu$ on the microwave power P. Equation (12) shows that the frequency bias is a function of the phase difference and the resonance linewidth. The linewidth, moreover, is a function of the microwave power level. This can be understood intuitively if one realizes that with decreasing microwave power the slower molecules become relatively more effective which results in a corresponding line-narrowing.[3] The frequency bias $\Delta\nu$ thus becomes power dependent. A numerical calculation of the relationship between linewidth and microwave power is depicted in Fig. 1. In Fig. 2 we depict the frequency bias $\Delta\nu$ as a function of both linewidth [eq (12)] and microwave power [eq (12) and Fig. 1]. The ordinate is normalized to $\Delta\nu/\Delta\nu_o = 1$ at optimum power, $P = P_o$ and $W = W_o$, where $\Delta\nu_o$ denotes the frequency bias at optimum power.

Equation (12) and Fig. 2 show that the power dependence of the frequency bias is almost linear; however, we note that the extrapolation of this near-linear portion to $P \to 0$ does not yield a vanishing frequency bias. A reduction of power from P_o to $P_o/2$ will result in a frequency bias change of about 1/3 of the total bias at optimum power or more accurately

$$\Delta\nu\,(P_o/2) \approx 0.70\,\Delta\nu_o. \quad (13)$$

V. Experimental Results

The reported measurements were performed during the accuracy evaluation of the cesium beam frequency standard, NBS-III, in 1969.[2] Figure 3 depicts the experimental results. Plotted is the fractional frequency change as a function of the microwave power level for opposing beam directions (solid lines). The precision of these measurements is discussed in Ref. 2. The microwave power changes ranged from P_o (optimum power) to not less than one-fourth P_o. The actual value for P_o was $P_o = 1.4$ mW; the frequency modulation amplitude was adjusted to $\nu_m \approx \nu_{mo}$. The two slopes are $+22 \times 10^{-13}$ per mW and -11×10^{-13} per mW. The total fractional frequency change with beam reversal at $P = P_o$ is 78×10^{-13} which gives us $\Delta\nu_o/\nu_o = 39 \times 10^{-13}$.

We believe that the asymmetry in the two slopes of Fig. 3 is caused by asymmetries in the spectrum of the microwave signal.* Such effects may be expected to be unchanged under beam reversal and to depend linearly on the microwave power (for a small single sideband perturbation).[3] We obtain symmetric slopes if we subtract a slope of $+5.5 \times 10^{-13}$ per mW. Thus we can conclude that the spectrum causes a power shift of $+5.5 \times 10^{-13}$ per mW (dashed line in Fig. 3), and the cavity phase shift a power dependence of $\pm 16.5 \times 10^{-13}$ per mW (dashed-dotted lines in Fig. 3).

As we expected theoretically, we have a seemingly linear functional dependence. The measured biases at optimum power and half-optimum power are respectively, $\Delta\nu_o/\nu_o = 39 \times 10^{-13}$ and $\Delta\nu(P_o/2)/\nu_o = 27.5 \times 10^{-13}$ or

$$\Delta\nu(P_o/2) \approx 0.71\ \Delta\nu_o \qquad (14)$$

which is in agreement with the theoretical results of Fig. 2 and eq (13).

We can now calculate the cavity phase shift ϕ from eq (12). The numerical calculation yields

$$\Delta\nu_o = 0.26\ W_o \phi\ . \qquad (15)$$

The linewidth was measured to be $W_o = 45$ Hz and we obtain from eq (15) with the power dependent frequency change discussed before $\phi \approx 3 \times 10^{-3}$ radians.

VI. Conclusions

We presented an analytical discussion of frequency shifts in a beam tube which relate to the presence of a cavity phase difference between the two Ramsey excitation regions The results are approximate and may be expected to be different for different experimental conditions (see Section II).*

The results, as summarized in eqs (10) and (12) and in Fig. 2 show that frequency shifts caused by cavity phase differences, the microwave power level, the resonance linewidth (velocity of atoms), and the frequency modulation amplitude are closely interrelated.

Frequency changes due to changes in the microwave power level are quite pronounced, which is contrary to the result of a previous treatment of this subject.[4] In fact, the power dependence could be utilized to determine the frequency bias when other causes for frequency bias are known to be absent, e.g., microwave spectrum effects (see Section V). In this case the measurement of the frequency change with reduction of the microwave power to, for example, $P_o/2$ yields directly the frequency bias according to eq (13). We have used this method in recent experiments on new cesium beam tubes.

A cavity phase difference will, in general, also cause frequency shifts to occur if the frequency modulation amplitude is changed. The frequency changes are virtually absent if a frequency modulation amplitude of less than half the linewidth is chosen (Section III). However, if larger frequency modulation amplitudes are used or if the cavity phase difference is unusually large (large bias) this effect could become significant. An analytical treatment which is valid for these conditions would require a higher order approximation than was attempted in this paper.

In the presence of a spectrum-related frequency bias of unknown magnitude, methods like the one discussed above will not be adequate. Beam reversal presently appears to be the only method** which then allows the separation and individual measurement of the different frequency biases. Beam reversal changes the sign in eq (12): consequently, one should then obtain microwave power (and frequency modulation amplitude) dependencies which are identical for the two beam directions, except for the sign reversal. Any deviation from this symmetry would indicate the presence of additional effects, e.g., signal spectrum asymmetry. The frequency biases can then be obtained as was demonstrated in Section V (Fig. 3).

*Other causes, e.g., the second-order Doppler effect which is totally ignored in this paper, must be taken into consideration when extreme accuracies (beyond the present state of the art) are the objective.

*The treatment of the power dependence due to cavity phase differences, as presented in this paper, leads to a rather simple physical picture: The chosen microwave power level acts like a velocity selector because only atoms within a rather narrow velocity range will have a significant transition probability. Atoms with velocities other than that selected by the given power level do contribute, but only in some minor fashion as may be seen from eq (12) and Figs. 1 and 2.

**Beam velocity changes,[5] as applied to the detection and correction of cavity phase differences, act in a way quite similar to the variation of the microwave power. Both act on W in eq (12). However, this method also should only be used with great caution if spectrum related frequency biases are present.

We summarize: In the absence of other biases, e.g., spectrum-related biases, variation of the microwave power offers a way to easily detect and correct for a cavity phase difference and the related frequency bias.

Acknowledgments

We thank Dr. Stephen Jarvis, Jr. for his assistance in computing the data of Fig. 1 and for his very valuable, constructive criticism.

References

1. Thirteenth General Conference of Weights and Measures (CGPM), Resolution 1 (1967).
2. D. J. Glaze, IEEE Trans. on Instrumentation and Measurement IM-19, p. 156 (1970).
3. N. F. Ramsey, Molecular Beams, Chapter V. Clarendon Press, Oxford (1956).
4. R. J. Harrach, J. Appl. Phys. 38, p. 1808 (1967).
5. G. Becker, B. Fischer, G. Kramer, and E. K. Müller, PTB-Mitteilungen 2, p. 77 (1969).

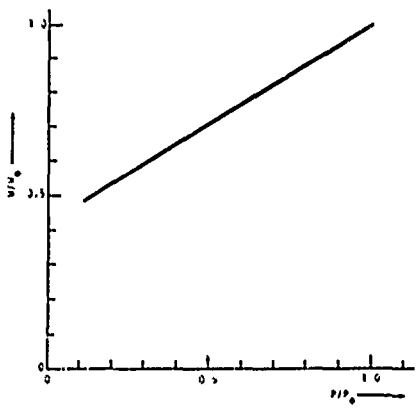


FIGURE 1 - Calculated dependence of the linewidth as a function of the microwave power. Linear approximation.

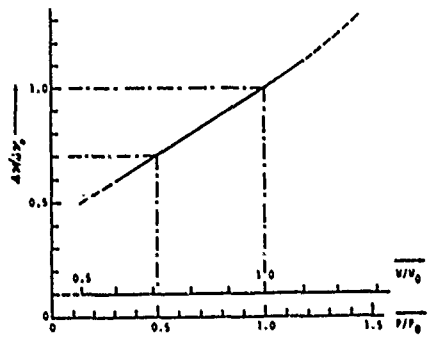


FIGURE 2 - Calculated Frequency bias (normalized) as a function of linewidth and microwave power.

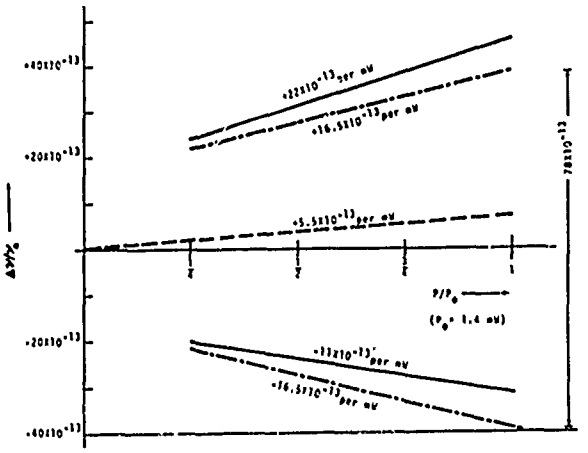


FIGURE 3 - Experimental results. Solid lines depict the fractional frequency shift measured as a function of the microwave power level (P_o = 1.4 mW) for both directions of the atomic beam. The dashed line represents the bias attributed to spectral impurities of the microwave signal. The dashed-dotted lines are obtained by subtracting this bias from the measured results and represent the bias due to a cavity phase difference. For details on the experimental procedure and a discussion of the measurement precision and accuracy, see Ref. 2.

PERFORMANCE OF NEWLY DEVELOPED CESIUM BEAM TUBES AND STANDARDS

R. Hyatt and D. Throne
Santa Clara Division
Hewlett-Packard Company
Santa Clara, California

and

L. S. Cutler, J. H. Holloway and L. F. Mueller
Hewlett-Packard Laboratories
Hewlett-Packard Company
Palo Alto, California

Summary

During the past few years, Hewlett-Packard has developed a new small (6.4") cesium beam tube, a new frequency standard which uses the small tube, and an improved 16" beam tube which can be retrofitted in existing HP cesium standards. These new products, developed on company funds, are designated the HP 5084 tube, the HP 5062A/B standard, and the HP 5083 tube, respectively. Part I of this paper deals with the development and characteristics of the new cesium tubes. Part II discusses the details of the 5062A/B standard and some aspects of the performance of a modified HP 5061 standard retrofitted with a new improved 16" tube.

Part I -- Beam Tubes

Introduction

Since the design principles and features in the two new beam tubes have much in common, the discussion of both tubes is carried out in parallel in this section. The discussion is divided into three main subdivisions:

1. Factors related to beam intensity.
2. Factors related to accuracy and stability.
3. Other design factors.

Finally, a table is presented to compare the important characteristics of the new tubes with the present 16" tube, the HP 6120. The data presented is based on measurements on a few prototype tubes and is expected to be typical for future tubes of this design. We do not yet have adequate data to make firm quantitative statements about the accuracy and reproducibility of these new products. More extensive measurements on a fairly large group of standards and tubes are required to give statistical significance to such specifications.

Factors Related to Beam Intensity

In the last fifteen years, considerable effort has been spent trying to arrive at an optimum design for the beam optics, i.e., the choice of deflection magnet characteristics, the choice of collimator and detector geometry, the division of the tube length between deflection and free drift regions, etc. The object is generally to achieve the maximum figure of merit[1] without using too much cesium. Considerable uncertainty has centered on the choice of deflection magnet, specifically, whether to choose a multipole magnet approach (using four or six pole magnets), dipole magnets (which have been used on all commercial cesium tubes by all companies to date), or perhaps a hybrid system. This is a difficult question to answer with accuracy. Even large computers are over-taxed by the large number of calculations required when simplifying approximations are not made. And, in some cases, the simplifying approximations have led to serious errors in the conclusions from such calculations. So, a combination of calculations and experiments, which generally are long and expensive, has been required to clarify this issue. On reviewing this information, we at HP decided to use dipole magnet optics, but with one innovation--we ran two beams in parallel in the improved 16" tube and 12 beams in parallel in the small 6.4" tube.* This decision was based on a comparison between a set of hexapole optics and double dipole optics for a 16" tube.** A hexapole tube was built whose figure of merit, corrected for mass spectrometer efficiency and multiplier noise, was measured to be about 9 $sec^{1/2}$ at an effective oven temperature of 110°C with a rate of cesium expenditure of 0.154 gm/yr. The computed signal for this tube would lead to a higher figure of merit under the approximations used in the calculation. Background signal is difficult to compute, but the actual background level is always high in multipole tubes, degrading the figure of merit. The calculations for hexapole tubes are not as satisfactory as those for dipole tubes. The double dipole optics calculations predicted a figure of merit of 15 $sec^{1/2}$ with an effective oven temperature of 110°C and an expenditure rate of cesium of 0.7 gm/yr. Constructed tubes give close to this performance.

The hexapole tube is thus somewhat more efficient in the use of cesium. This efficiency is given by:

$$Q = F^2 R ,$$

where F is the figure of merit and R is the cesium expenditure rate. For the hexapole design we get

*Patent applied for.

**The work on the 16" multipole tube was supported by the Naval Ship Systems Command, Contract No. N0024-67-C-1048.

Q = 530 year sec/gm and for the double dipole Q = 320 year sec/gm. However, the dipole design is capable of using the cesium at a greater rate without self-scattering in the beam because the oven apertures can be larger. This allows a figure of merit larger by a factor of about 1.67. In addition, multipole magnets are more expensive to make and do not lend themselves to a multiple beam configuration.

Figure 1 shows the magnet pole piece used on the small tube. Note that while there is an array of parallel beam-deflecting gaps, the structure is not a multipole magnet. The motivation for using parallel beams was primarily to increase the beam intensity, but it was soon recognized that an important extra benefit was available--if the beams are stacked back to back (as shown in Figure 1), greatly improved performance under acceleration is realized. Suppose an acceleration is applied in the direction of the beam deflection (normally the most sensitive direction to acceleration). The signal from one set of beams will decrease, but this decrease tends to be balanced out by an increase in the signal from the other set of beams. So there is a great reduction in the net signal change under acceleration. This has been checked by measuring the tube output signal while rotating the tube. The results are shown in Figure 2.

Factors Related to Accuracy and Stability

The microwave cavity and the "C" field (i.e., the uniform field of the order of 0.06 gauss in the interaction region) are the most important factors in determining accuracy and stability of the frequency which characterizes a given beam tube. These elements are discussed in this section.

Both new tubes use a conventional "U"-shaped Ramsey cavity which is milled in two mating sections from solid blocks of metal. This manufacturing method is consistent with the high mechanical accuracy and stability requirements of the structure. The characteristics of these structures are summarized in Table I. Measurements have been made on the electrical length asymmetry of the arms of the cavity. The cavity phase difference produced by the measured asymmetry would cause frequency shifts no larger than 2×10^{-13} in the improved 16" tube and no larger than 7×10^{-13} in the small tube.

The magnetic shields in both new tubes are similar. Both are triple shielded. The innermost shields are rectangular boxes, closed on all sides, with apertures for the beam and the microwave input guide. The boxes contain exciting coils to generate the C field. Since the aperture for the input guide is relatively large, a "baffle" shield with an exciting winding is placed between it and the beam to reduce the perturbation from the aperture. The combination of these design features with an improved degaussing method* has led to a large improvement in the shielding factor and in C field stability. The new degaussing technique involves the use of a moderately small, very low frequency decaying current. The tube is initially degaussed after the standard has been placed in operation and the C field turned on. After this, a very low level degauss can be applied while the standard is being used as a clock with negligible perturbation to its operation. This is very useful when it is desired to make small, deliberate changes in the C field. This procedure should reduce C field drift to very small values. The degausser can easily be built into the equipment if desired. If the shields are deliberately magnetized so as to shift the Zeeman frequency about 2 kHz, they can be degaussed reproducibly within about 5 Hz, corresponding to frequency offsets of less than 1×10^{-13}.

The results of an experiment to determine the shielding of an improved 16" tube are shown in Figure 3. Note that the shielding factor measured here is about one million. The effect of various orientations of the earth's field should be only a few parts in 10^{14}.

A test was run on the temperature stability of the improved 16" C field. A temperature change of 30°C produced a maximum shift of 0.04% in Zeeman frequency which corresponds to a fractional frequency change of about 1.4×10^{-13}.

The microwave spectra of the new small and improved tubes are shown in Figures 4 and 5, respectively.

Other Engineering Features

The detector section of the tubes received a considerable amount of attention. A new, ruggedized ionizer was designed, similar in both tubes. The lowest mechanical resonance occurs at about 900 Hz for the small tube and at about 1100 Hz for the 16" tube. Tests showed no observable frequency shifts when the tubes were vibrated at or near the resonance frequencies.

The mass spectrometers have approximately 50% efficiency and good rejection for potassium. They were designed with the help of potential models using resistance paper and conductive paint to make two-dimensional models and an electrolytic tank for the cases requiring cylindrical symmetry.

The electron multipliers in both tubes require only six stages since the beam currents are fairly high due to the efficient design and multiple beams.

Careful attention has been paid everywhere to make the tubes extremely rugged and reliable. The experience gained from the large number of tubes of the earlier design in operation in the field has been of great value. Particular efforts have been made to raise the mechanical resonance frequencies as high as possible. Adequate cesium and cesium getters have been supplied to give an estimated life of 5 years for continuous operation.

Figure 6 shows the new small tube compared to a standard 16" tube.

*Patent applied for.

Part II -- 5062A/B Cesium Standard

Introduction

The development of the miniature cesium beam tube has opened up a new area of application for precision time and frequency standards. The cesium standards presently available were designed primarily for laboratory use whereas the small ruggedized tube has made possible the design of a compact primary frequency standard for field and systems use.

The HP 5062A/B Cesium Standard has been developed to fill the need for precision time and frequency in such systems as the Aircraft Collision Avoidance System as presently defined by the Air Transport Association.[3] Figure 7 shows a size comparison between the airborne version of the 5062A/B and the HP 5061A Cesium Beam Frequency Standard. The 5062A/B is about half the weight and one-third the volume of the 5061A.

The protrusion on the front panel of the 5062B is a detachable battery pack which is present in the 5062B commercial airline version only. The only input-output to the 5062B is through a special ARINC rack connector located on the rear panel.

The front panel of the 5062A, which is a commercial version of the 5062B, will have logic indicators, a magnetic field control and some circuit monitoring functions. The rear panel will have conventional power plugs and output connectors.

Another possible configuration would be a 16-3/4" wide by 5-1/4" high by 16-1/2" deep rack package. This is the same package size as the HP 5065A Rubidium Standard. The circuitry and beam tube required for the cesium standard would occupy about one-half of the volume of this package, leaving the remaining half available for user options such as clocks and special output frequencies.

Operational Considerations

To fulfill the requirements of a ruggedized standard, the following environmental design considerations were established.

The unit should operate continuously from -55°C to +74°C. All circuitry has been designed to operate over this temperature range. The design goal on frequency offset over this temperature range is less than 1×10^{-11}.

The standard should operate under vibration from 0 to 55 Hz, 0.01 inches double amplitude displacement and 55 Hz to 2000 Hz at 1G level. The frequency offset should be less than 1×10^{-11}.

The magnetic field susceptability should be less than 1×10^{-13} for magnetic fields of less than 1 gauss uniform dc or 1 gauss peak ac.

The unit should withstand MIL-STD-901C shock and should survive storage from -54°C to +85°C for a period of up to two weeks with no detrimental effects.

Design Considerations

The 5062A/B control system is basically the same as that used in the HP 5060A Cesium Standard and was described in detail by Cutler and Bagley in 1964.[3]

The 5 MHz slave oscillator uses a high temperature bake-out crystal to achieve low aging, good retrace and fast warmup. The aging rate of less than 5×10^{-10} per day and electronic control range of $\pm 1.5 \times 10^{-7}$ provides a minimum of 300 days operation with no adjustment required. Since typical oscillators achieve aging rates of about 1 to 2×10^{-10} per day after one to two weeks of continuous operation, the oscillator adjustment interval could be extended to nearly four years.

The oscillator uses a single stage oven to achieve a temperature coefficient of 5×10^{-9} from -55°C to +74°C. Fast warmup to 2.4×10^{-9} in fifteen minutes is achieved with an input power of 15 watts.

Good spectral purity is achieved with low noise feedback amplifiers and a crystal filter with a 50 Hz bandwidth at the output of the oscillator.

The oscillator has been vibration tested with good results. No resonances were evident in the range from 55 Hz to 2500 Hz at 1g vibration. The oscillator maintains spectral purity with vibration sidebands down -90 dB at 55 Hz and greater than -100 dB at frequencies from 500 Hz to 2500 Hz.

The 5 - 90 MHz multiplier supplies 400 mW of RF power to drive a step recovery diode. The step recovery diode is used to generate 9180 MHz and is also used as a mixer to mix 12.63 . . . MHz to obtain the cesium transition frequency of 9192.63 . . . MHz. The output of the 90 MHz multiplier varies less than ± 0.5 dB from -55°C to +100°C. This produces about ± 0.5 dB variation in the output of the step recovery diode at 9.180 GHz.

The step recovery diode assembly is temperature controlled to meet the wide operating temperature range.

The synthesizer uses a preset digital divider to phase lock a VCXO to the 5 MHz signal from the quartz oscillator. A sampling phase lock loop is used to lock to a harmonic of the frequency out of the digital divider. The sampling sidebands are kept below -80 dB.

Switching regulators are used exclusively in the 5062A/B to maintain high efficiency and low temperature rise inside the instrument. The power supplies consist of a +18 volt 600 mA main power regulator, -3500 Vaclon supply, -2500 electron multiplier supply, -18 volt bias supply, +5 volts at 500 mA and +1 volt at 2 amps for the hot wire ionizer. The overall power supply efficiency is about 60%.

The remaining circuitry consists of a preamp and tuned amplifier to amplify the error signals from the beam tube and a reference oscillator and phase detector to translate the ac error signal to dc. An integrator with open loop dc gain of 100,000 is used to set the noise bandwidth of the servo loop.

The design goals require that individual electronic effects contribute less than 1×10^{-12} to the frequency offset over the temperature range -55°C to +74°C.

The instrument has automatic lockup circuitry which prevents lock on a sidelobe of the Ramsey pattern. The scheme utilizes the good retrace and fast warmup characteristics of the slave oscillator. The broad line width of the small tube (approximately 1300 Hz) allows lock up if the crystal oscillator has a control range of $\pm 1.4 \times 10^{-7}$ (1300 Hz at 9.2 GHz) and will retrace to within $\pm 1.4 \times 10^{-7}$ of the correct frequency. The logic circuits detect the presence of second harmonic, in phase fundamental or quadrature phase fundamental of the modulation signal. A proper locked condition is indicated by the absence of fundamental and the presence of second harmonic of the modulation signal. If this condition is not met, the logic circuit toggles the servo loop open and closed at a 15 second rate until lockup is achieved. When the oscillator drifts to within 1.5×10^{-7} of the correct frequency, lockup will occur.

There are dual 5 MHz outputs with -80 dB isolation and -40 dB distortion. Variation of load in one output from open to short will cause less than 2 nanoseconds phase shift and less than 1% amplitude variation on the other output.

Reliability

The reliability of any electronics equipment is directly related to the number of components contained in the instrument and the operating temperature of the components. The 5062A/B contains 95 transistors, 27 integrated circuits, 369 resistors, 212 capacitors, 50 diodes and 40 inductors for a total of 793 components. This represent about two-thirds the number of parts contained in the HP 5061A Cesium Standard. Therefore its reliability should be even greater than observed for the 5061A Cesium Standard.

A theoretical MTBF of 14,000 hours has been calculated using HP corporate failure rates for components. This compares with a calculated failure rate of 9300 hours for the HP 5061A Cesium Standard. In actual practice, the 5061A has exhibited an MTBF in excess of 18,000 hours, as stated above.

At this point it is impossible to give reliability figures that are really meaningful for the small cesium beam tube since it has no history. However, the materials used and many of the construction techniques are the same as those used in the 16" tube presently used in the HP 5061A Cesium Standard. The MTBF of that tube is about 150,000 hours as calculated from actual field data on over 100 tubes. Such things as flux into cesium getters do not scale directly but this has been taken into account in the design. Therefore, extremely good reliability is expected from the small tube. It should be noted that the MTBF of more than 15 years for the present 16" tube indicates that the majority of beam tube failures will be due to end-of-life mechanisms rather than catastrophic failure. This should remain true with the small tube.

The failure rates for field use will obviously be higher due to the harsh environment. For example, we would expect electronic failures to reduce the calculated MTBF to about 6,000 hours in commercial aircraft application.

All electronic assemblies exclusive of the quartz oscillator, harmonic generator and beam tube have been designed to operate continuously at 105°C ambient temperature. All components are dissipating less than 75% of rated power at 105°C. This design margin insures very reliable operation at temperatures up to 74°C.

The temperature rise in the instrument at 25°C ambient varies from 5°C to 10°C at the hottest spot.

Maintenance

The package design and extensive use of plug-in printed circuit boards makes troubleshooting and repair of the instrument very easy. Figure 8 shows printed circuit boards pulled out of the front and rear card cages. The boards in these two card cages plug into mother boards shown in Figure 9. The interconnections between mother boards is made with conventional cabling and connectors.

The forward portion of the instrument contains the clock option, servo loop, logic and quartz oscillators. The center section contains the beam tube and frequency multipliers. The rear section is devoted to power supplies and output amplifiers.

There are a total of twenty-five test points identified in the instrument and brought to a rear test connector in the commercial airline package. A troubleshooting tree used with these test points will allow a technician to isolate a faulty module and make a speedy repair.

Performance Tests

Initial temperature testing of the first prototype over the temperature range -25°C to 74°C gave a frequency shift of -2×10^{-12} at both temperature extremes. The power required at +25°C was 25 watts and at -25°C was 28 watts.

Warmup tests were conducted and a warmup time of 15 minutes to achieve rated stability was observed. The instrument requires 120 watts of warmup power for less than 10 minutes.

ac and dc magnetic field testing was carried out on the first prototypes. The instrument was placed inside a Helmholtz coil and subjected to 3 gauss rms at 60 Hz. No offset was observed in two planes to a resolution of 1×10^{-13}. In the third plane an offset of 4×10^{-13} per 1 gauss rms was observed.

The effect of dc magnetic field on the field independent line was too small to be easily measured directly. The measurement was made by locking the instrument on the first magnetic field dependent line and measuring the effect on the Zeeman transition. The largest shift observed in any direction for fields of 5 gauss was 30 Hz. For Zeeman offsets of 80 kHz, the effect on the field independent transition is about 5×10^{-13}.

Vibration testing has been started and the instrument passed a low frequency test of 0.01 inches double amplitude displacement 0 to 65 Hz in all planes. A phase step of less than 2 nsec was observed at about 46 Hz. Dwelling at this frequency produced no larger discontinuities. No package resonances were observed over this frequency range.

Short-Term Stability

The stability of frequency standards which are perturbed by various spectral densities of noise has been treated in detail by Cutler and Searle[4] and by Allan.[5] Figure 10 shows the expected value of the variance of the frequency fluctuations versus averaging time for the 5062A/B Cesium Standard. The data for averaging times from 10^{-4} seconds to 90 seconds were obtained with the beat frequency measurement system described by Cutler and Searle.[4] The data were analyzed on an N = 2 basis with a 100 kHz bandwidth as described by Allan.[5] The data at 3600 seconds was taken from continuous phase records using the method described by Allan. Each data point represents 100 measurements of sample size 2.

The data from 10^{-4} sec to 90 seconds were obtained with an HP 5360A Computing Counter which was programmed to calculate the value of the frequency fluctuations.

The frequency stability for averaging times less than the closed loop time constant of 100 msec is determined by the slave crystal oscillator. For averaging times longer than 200 msec the stability is determined by the cesium beam tube reference. For averaging times much greater than the closed loop time constant, the frequency stability averages with a slope $\tau^{-1/2}$. We see that this law is followed out to 3600 seconds where the fractional stability is 9×10^{-13}.

The spectral purity of the 5 MHz output of the 5062A/B is shown in Figure 11. These spectra were obtained by multiplying a reference oscillator and the 5062A/B output up to 9180 MHz and 9200 MHz, respectively. These two frequencies were mixed and the resulting 20 MHz difference frequency amplified and displayed on an HP 8552A Spectrum Analyzer.

Figure 11(a) shows a noise floor of about -64 dB with 100 Hz bandwidth on the spectrum analyzer. There are no spurious signals above -60 dB out to 100 kHz from the carrier.

Figure 11(b) shows the spectrum out to 25 kHz from the carrier with 100 Hz bandwidth. The same noise floor is evident and no spurious signals are detected above -60 dB.

Figure 11(c) shows the spectrum out to 5 kHz from the carrier with 50 Hz bandwidth. Spurious signals are down about -60 dB. Taking into account the 65 dB multiplying factor, spurious signals from 500 Hz to 100 kHz from the carrier are more than -125 dB down at 5 MHz.

Summary

A compact rugged cesium standard suitable for field and systems use has been described. Data has been presented which shows performance comparable to present commercially available cesium standards such as the HP 5061A. The major contribution of this standard will be its ability to satisfy these performance requirements under adverse environmental conditions.

The design of the instrument has incorporated many cost-saving and maintenance-oriented features such as plug-in printed circuit boards and numerous test points.

These factors, along with the inherent long-term stability and fast warmup of the cesium beam tube, should make the 5062A/B an excellent candidate for critical timing applications.

A summary of specifications is shown in Table II.

Performance of Improved 16" Tube in 5061 Electronics

Some performance tests have been made on the improved 16" tube. The results of some of these are contained in Table III.

The static magnetic field tests are summarized in Figure 3. An ac magnetic field test was run with an improved 16" tube mounted in a modified set of 5061 cesium standard electronics. The unit showed less than 1×10^{-13} frequency change for any orientation in a 2 oersted peak 60 Hz field.

A few measurements were made on frequency stability using two instruments fitted with the improved tubes. Averages of 240 seconds gave an Allan Variance $\sigma(2,240\ \text{sec})$ of about 4.4×10^{-13} for each standard. The tubes used both had a figure of merit of about 15, which would give a predicted performance of about 3.7×10^{-13} for 240 seconds. Thus the realized performance is close to that predicted. It is important to note that an improvement by a factor of 10 in figure of merit results in a reduction of time required to achieve a given precision by a

factor of 100 for times greater than the servo loop time constant.

Conclusion and Acknowledgment

We have described two new cesium beam tubes and a new small cesium standard. The tubes have greatly improved performance as a result of careful beam optics design, multiple beams, better magnetic shielding and C field structures, and precision-machined microwave cavities. The new small standard was designed to be very compact, rugged and reliable, and to operate and give excellent performance in severe environment.

The authors are pleased to acknowledge valuable contributions from Armin Baur, Lee Bodily, Jack Elmberg, James Koch, Richard Lacey, Gary Seavey, Hans Van Heyst, and Jim Zellers. Robert Kern, no longer with Hewlett-Packard, was originally in charge of the improved 16" tube project and made many valuable and lasting contributions to its design. The study of multipole optics referenced in this paper was supported by the Naval Ship Systems Command, Contract No. N00024-67-C-1048. The support given over the years by the Navy, Air Force, and Army to the many useful studies on beam tubes and their optics has indirectly benefitted our new tube program.

References

1. R. F. Lacey, A. L. Helgesson, J. H. Holloway, "Short-Term Stability of Passive Atomic Frequency Standards," Proc. IEEE 54, 2, 170 (Feb. 1966).

2. "Air Transport Time-Frequency Collision Avoidance System," Aeronautical Radio, Inc., ARINC Characteristic No. 587, September 1, 1970.

3. A. S. Bagley, L. S. Cutler, "A Modern Solid-State Portable Cesium Beam Frequency Standard," Proc. 18th Annual Symposium on Frequency Control, p. 344.

4. L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," Proc. IEEE 54, 2, 137 (Feb. 1966).

5. D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. IEEE 54, 2, 221 (Feb. 1966).

TABLE I

	Material	Drift Length	Interaction Length	Linewidth (Hz)
Present 16" Tube	Monel (silver clad)	4.4"	0.4"	550
Improved 16" Tube	Aluminum	6.6"	0.4"	365
6" Tube	OFHC Copper	1.8"	0.2"	1300

VITAL STATISTICS OF THE BEAM TUBE CAVATIES

TABLE II

5062A/B Cesium Standard Tentative Specifications

Accuracy:	$\pm 2 \times 10^{-11}$
Reproducibility:	$\pm 1 \times 10^{-11}$
Long-term Stability:	$< \pm 1 \times 10^{-11}$ for life of beam tube
Short-term Stability:	$< 5 \times 10^{-9}$ rms/100 μsec
Beam Tube Figure of Merit:	1.5
Warmup Time:	< 15 minutes from -15°C
Output:	5 MHz sine wave, 1 volt 50Ω
Harmonic Distortion:	< -40 dB
Non-Harmonic:	< -100 dB
Time Scale:	Adjustable from Atomic to UTC offsets of -700×10^{-10}
Environmental:	
Temperature:	MIL-E-5400H, Class 1A
Operating:	-55°C to +71°C $< 1 \times 10^{-11}$
Storage:	-62°C to +80°C
Humidity:	95% @ 50°C, $< 1 \times 10^{11}$
Altitude:	0 to 50,000 feet $< 1 \times 10^{-12}$
Vibration:	MIL-E-5400L (Curve IIA) $< 2 \times 10^{-11}$
Shock:	MIL-STD-901C
EMI:	MIL-STD-461
Magnetic Field (static and ac):	0 - 2 gauss $< 2 \times 10^{-12}$
Power:	28 watts 22 to 32 volts @ 25°C
Size:	4-7/8" × 7-5/8" × 19-9/16" 16-3/4" × 5-1/4" × 16-1/2"
Weight:	35 lbs.

TABLE III

Beam Tube Comparison Using
Typical Experimental Results

	In Production	New	
Characteristics \ Tube	6120 (standard 16")	5083 (improved 16")	5084 Small Tube
LENGTH	16"	16"	6.4"
DIAMETER	5.6"	5.6"	4.2"
VOLUME	330 in^3	330 in^3	95 in^3
WEIGHT	17 lb.	23 lb.	11 lb.
POWER (25°C ambient)	8 watts	6 watts	4 watts
ACCURACY	$\pm 2 \times 10^{-12}$	$\pm 2 \times 10^{-12}$	$\pm 5 \times 10^{-12}$
SHORT-TERM STABILITY (1 sec. avg.)	5×10^{-11}	6×10^{-12}	4×10^{-11}
MEASUREMENT TIME FOR 1×10^{-12} PRECISION (1σ)	45 minutes	36 seconds	30 minutes
RELATIVE SENSITIVITY TO ACCELERATION	1	0.2	0.07
SENSITIVITY TO STATIC MAGNETIC FIELD	1×10^{-12} pk-pk for earth's field	1×10^{-13} pk-pk for 5 Oe pk-pk	5×10^{-13} pk-pk for 5 Oe pk-pk
LIFE (continuous operation)	4 years	(5 years)*	(5 years)*

( )* not experimentally verified.

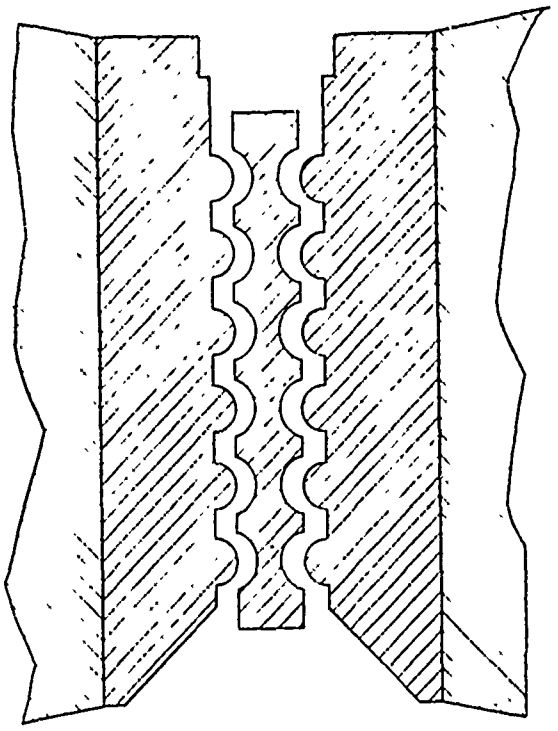
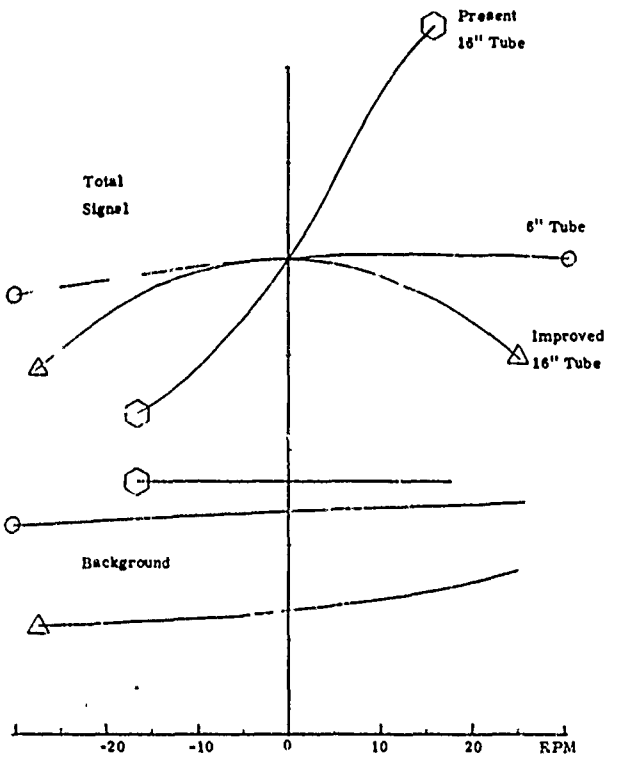
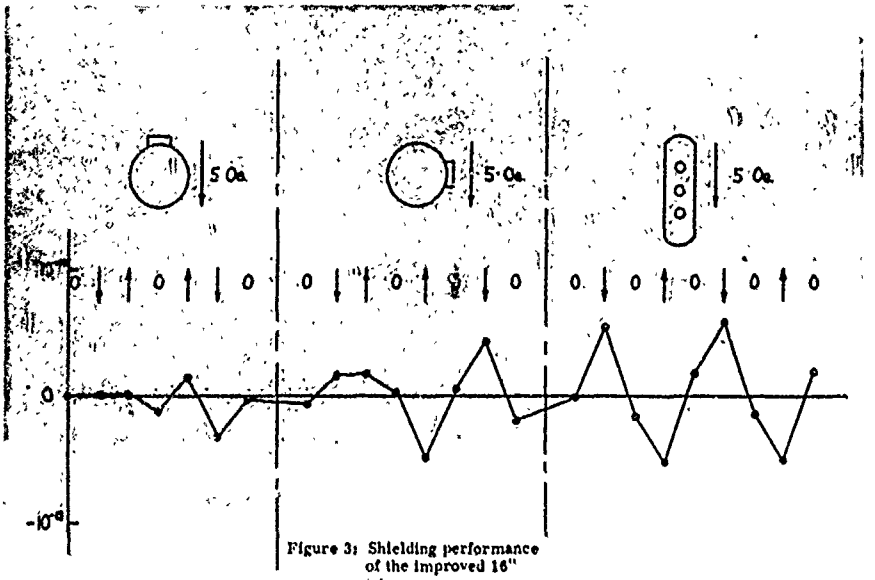


Figure 1: Multiple beam pole piece.

Figure 2: Curves of beam current versus rotation rate showing acceleration performance.




5 Oe
5 Oe
5 Oe
0
0
0

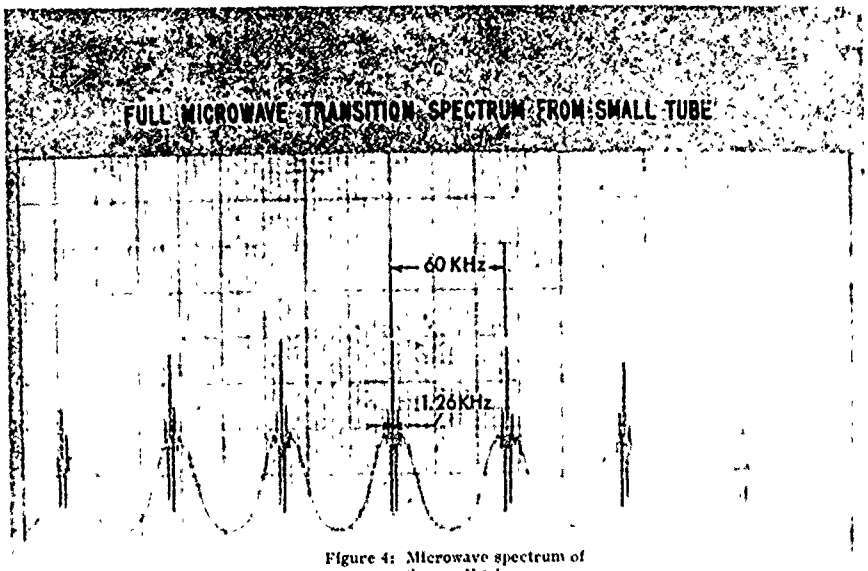


Figure 4: Microwave spectrum of the small tube.

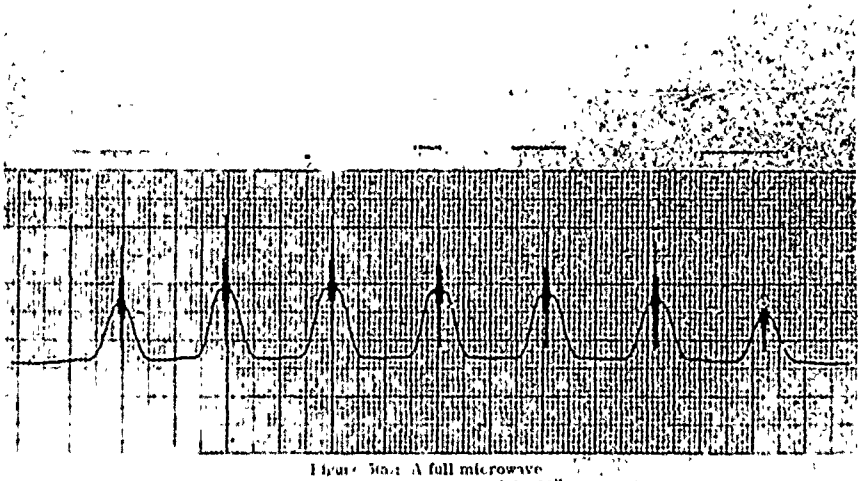


Figure 5a: A full microwave spectrum of the 16" tube.

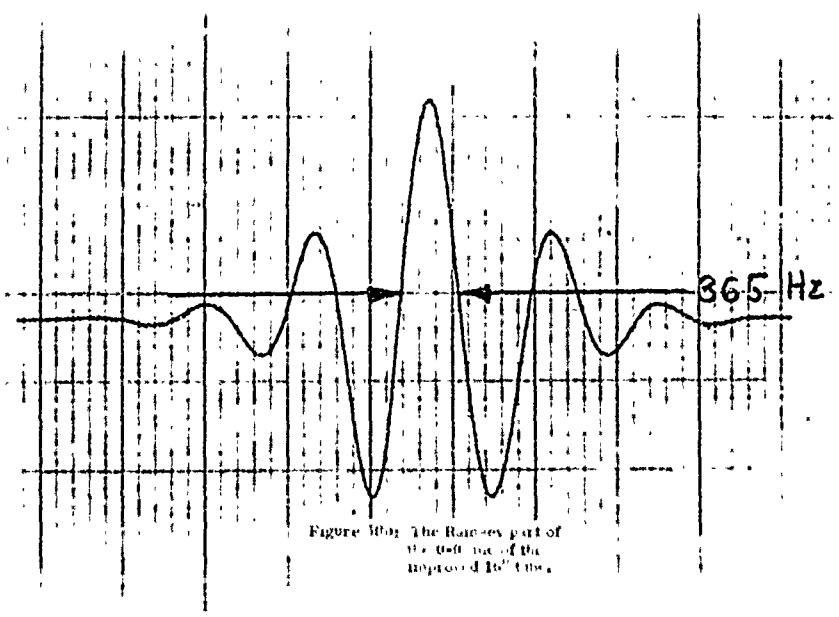


Figure 5b: The Ramsey part of the [illegible] of the improved 16" tube.

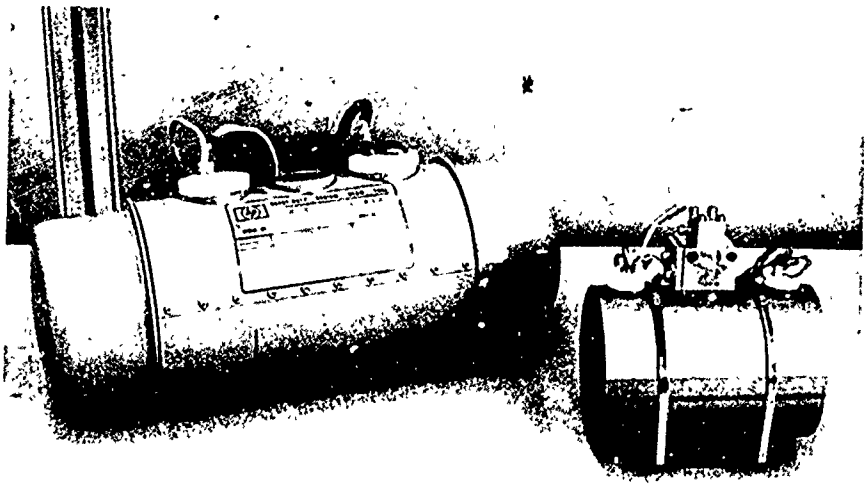


Figure 6: The new small tube compared against a standard 16" tube.

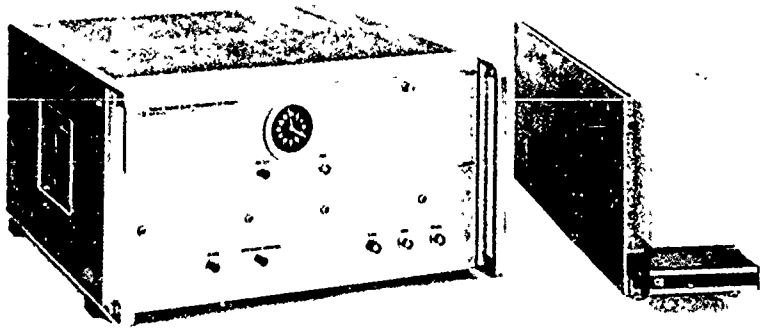


Figure 7: A comparison of the size of the 500A Certain Standard to the 401A Certain Standard.

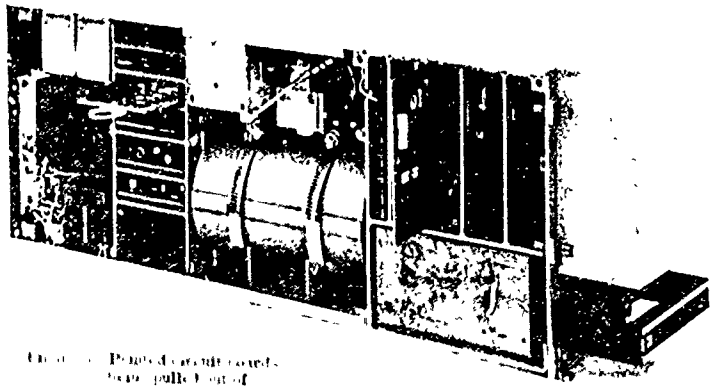


Figure 8: Power drawer unit can be pulled out of cabinet as the 401A is used.

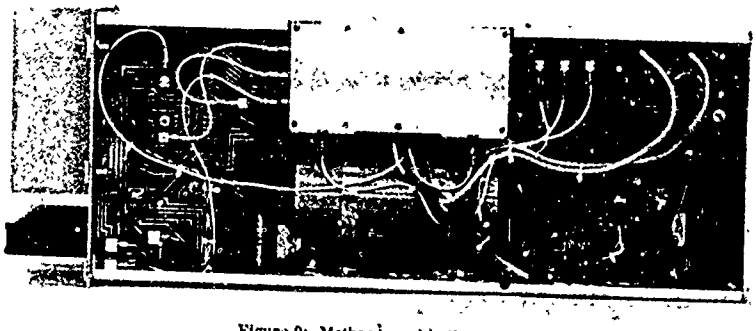
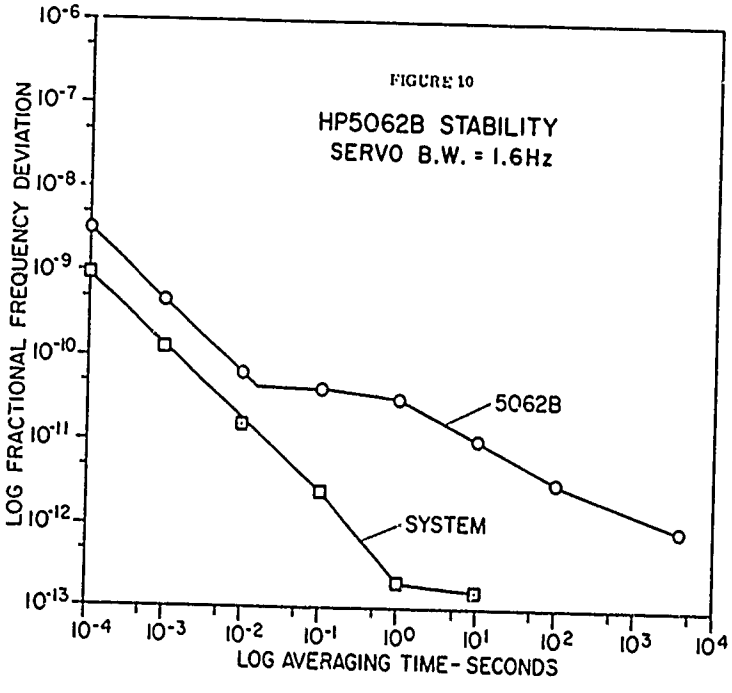
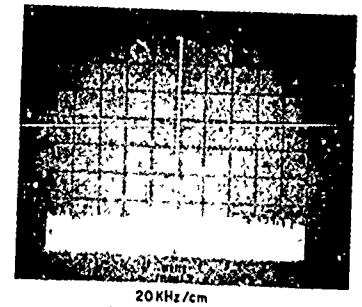
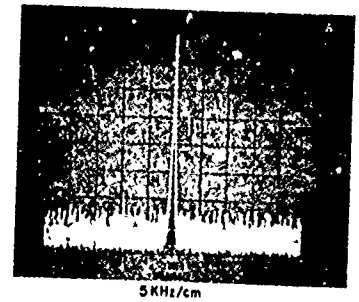


Figure 9: Mother board in the 5062A. Note the new small tube.







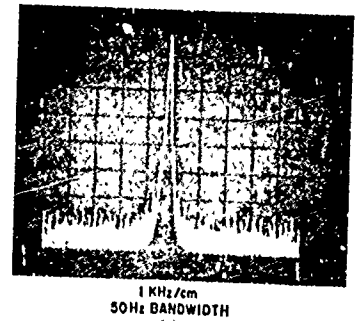


FIGURE 11 5062B SPECTRUM AT 9.2 GHz

A FIELD INDEPENDENT OPTICALLY PUMPED ^{85}Rb MASER FREQUENCY STANDARD*

W. A. Stern
General Time Corp.
Research Center
Stamford, Connecticut 06902

and

R. Novick
Department of Physics
Columbia University
New York, New York 10027

Summary

Continuous self-sustained oscillation at 3035.73...MHZ between the field-independent ground state hyperfine levels of ^{85}Rb, $5^2S_{1/2}$ (F=3, m_f=0) to $5^2S_{1/2}$ (F=2, m_f=0), has been obtained in a magnetically unshielded optically pumped maser frequency standard.

Operation of the maser is analyzed in terms of power output, the oscillation parameter and the pumping rate.

Light shifts due to both real and virtual transitions are analyzed.

* * *

At 3035.73...MHz continuous self-sustained oscillation between the field-independent ground-state hyperfine levels of ^{85}Rb, $5^2S1/2$ (F=3, m_f=0) to $5^2S_{1/2}$ (F=2, m_f=0), has been obtained in a magnetically unshielded optically pumped maser.

The maser oscillator described here operates on the field-independent transition in fields as high as 0.8 G with inhomogeneities of up to 0.2 G. This relative insensitivity to magnetic disturbances together with an observed power output of 8×10^{-10} W gives this maser oscillator good short-term frequency characteristics. Simple estimates indicate a stability of better than 1 part in 10^{13} for averaging times between 1.0 and 0.01 sec. Measurements of the relative phase stability of the ^{85}Rb maser are underway and will be reported later. The ^{85}Rb maser will be limited in its long-term stability by cavity pulling, changes in the pumping lamp profile, as well as chemical changes in the buffer gas due to outgassing. However, several schemes have been proposed to solve these problems.

A large population inversion between the $5^2S_{1/2}$ (F=3, m_f=0) and $5^2S_{1/2}$ (F=2, m_f=0) levels of ^{85}Rb can be efficiently obtained by intensity pumping. Intensity pumping is possible because of the fortuitous conincidence of certain hyperfine transitions in the ^{85}Rb-^{87}Rb system. A 1 1/4-in.-diam. resonance lamp containing 99.85% pure ^{85}Rb and 1.8 Torr of argon is excited by a 20-MHz rf discharge. The ^{85}Rb resonance radiation, containing both the $5^2S_{1/2}$, F=3 - 5^2P and $5^2S_{1/2}$, F=2-5^2P resonance lines, is passed through a filter cell filled with ^{87}Rb at 80°C helium at a pressure of 70 Torr. With sufficient buffer gas pressure broadening and shifting, the $5^2S_{1/2}$, F=2 - 5^2P ^{87}Rb absorption line can be made to almost completely absorb the unwanted $5^2S_{1/2}$, F=3-5^2P ^{85}Rb emission line normally present in the ^{85}Rb resonance lamp Since the filtered resonance radiation predominately contains the $5^2S_{1/2}$, F=2 - 5^2P resonance line, it causes overpopulation in the upper $5^2S_{1/2}$, F=3 level in the ensemble of ^{85}Rb atoms. In the steady state, assuming that the $5^2S_{1/2}$, F=2 ^{85}Rb level is completely depopulated, each $5^2S_{1/2}$, F=3 magnetic substate has a relative population density of $\frac{1}{7}$. In ^{87}Rb, the $5^2S_{1/2}$, F=2 level has five magnetic sublevels. Hence the number of atoms available for maser action in ^{85}Rb is $\frac{5}{7}$ of those available in ^{87}Rb under similar conditions.

The pumped ^{85}Rb is contained in a vacuum-tight TE_{021} microwave cavity resonant at 3035.73...MHz and filled with nitrogen gas at a pressure of 8 Torr. The nitrogen buffer gas reduces the collision frequency of the aligned ^{85}Rb atoms with the walls, and quenches reradiation from the upper P states. The cavity is made of copper-plated 305 nonmagnetic stainless steel. Pumping light is admitted through Pyrex windows sealed to the ends of the cavity with nonmagnetic stainless-steel housekeeper seals. The light enters the cavity region through perforated end walls. Tuning is accomplished over a 2-MHz range by a tuning stub. A movable microwave loop provides proper coupling to the desired mode After evacuation, 25 mg of

^{85}Rb metal (99.85% pure) are introduced into the cavity and settle in a pool on the bottom of the cavity. Very severe ^{85}Rb density gradients appear initially and inhibit operation of the maser until the ^{85}Rb is uniformly distributed throughout the cavity. This is accomplished by heating the cavity locally until the metal is dispersed throughout the cavity and uniformly coats the walls. After proper mixing with the buffer gas the ^{85}Rb density gradients disappear.

The unloaded cavity Q is 85 000 and is high enough to permit oscillation with a relatively small cavity-filling factor and pumping light intensity. In addition to the high cavity Q, the large cavity volume (9×10^3 cm^3) permits high power output for low vapor densities and facilitates the study of various relaxation processes within the vapor.

The maser is operated at 50 °C and is stabilized to within 0.01 °C by a simple temperature control system. The nitrogen buffer gas pressure is not critical and the maser has been operated over a range of pressures from 6 to 12 Torr with little change in output power. Insufficient light intensity was available to quench oscillation. The maser oscillates over a 2:1 range of light intensity and the corresponding change in frequency (light shift) is 10Hz

The maser output is detected with a standard 30 MHz, i.f. microwave superheterodyne receiver. Figure 1 shows the maser signal at the output of the i.f. strip during cw operation. Figure 2 shows the maser output at the same point when the pumping light is turned on and off. At t=T1 the light is turned on and the maser signal rises exponentially from the noise level until cw operation begins. At t=T2 the pumping light is shut off. At first the maser output increases because the disorienting effect of the pumping light is removed. This lasts for a few milliseconds before the output decays exponentially to zero

It is extremely useful to describe the operation of the maser as function of lamp pumping rate, relaxation times, temperature, filling factor, and cavity Q. Γ'_m is designated the oscillation parameter because it has an upper limit for which oscillations are not possible. By requiring $\frac{P}{P_m} > 0$ it is found that for self-sustained oscillations occur when :

$$\Gamma'_m < \frac{\Gamma'(1+\Gamma')}{\frac{7}{2}\Gamma'^3 + (\frac{19}{2} + 7r)\Gamma'^2 + (6 + 19r)\Gamma' + 12r}$$

where $p_m = \frac{1}{2} N h\nu \Gamma_m$

$r = \gamma_2 / \gamma_1$

$\Gamma'_m = \Gamma_m / \gamma_1$

$\Gamma' = \Gamma / \gamma_1$

$\Gamma'_m = \gamma_1 \hbar / 4n\pi Q \eta \mu_0^2$

Q = Cavity Q

η = filling factor

n = ^{85}Rb density

N = total number of atoms

V_c = Volume of cavity

ν = cavity resonance frequency

μ_o = Bohr Magneton

For low temperatures $\approx$ 45 °C where spin exchange relaxation is low, $r > 1$. At higher operating temperatures $\approx$ 70 °C it is found that $r \approx 1$ because spin exchange becomes a dominant relaxation mechanism. The oscillation parameter as a function of pumping rate is shown in Figure 3 for the case r = 1.

In order to obtain oscillation it can be seen from the equations for the oscillation parameter Γ'_m that there are certain constraints on the power output, cavity Q, filling factor and operating temperature of the ^{85}Rb maser. For example, a maser using a high mode, high Q cavity will not oscillate at high temperatures because of the absorption of the pumping light in the first few centimeters of the cavity. On the other hand a maser using a lower mode, lower Q cavity will only oscillate at higher temperatures where the higher ^{85}Rb density compensates for the low Q. Table 1 lists the minimum value of the oscillation parameter required for oscillation for various values of $r = \gamma_2/\gamma_1$.

TABLE 1

Minimum Value of the Oscillation Parameter Required for Oscillation.

Γ'_m	$r = \gamma_2/\gamma_1$
0.035	1

0.025	2
0.020	3
0.015	4

Note that at lower temperatures, where $r > 1$, the oscillation condition becomes harder to realize.

Figure 3 shows the quenching of oscillation due to too large a pumping rate Γ'. In the present maser the light intensity was insufficient to quench oscillation. However the experimental data shown in Figure 4 shows that enough light was present to cause saturation of the power.

One of the sources of phase instability in the ^{85}Rb maser is the shift in the ground state hyperfine frequency due to changes in the pumping light flux and filtered lamp profile. Arditi and Carver (1) presented the first experimental evidence of such shifts for the field independent transitions of ^{133}Cs and ^{87}Rb. Barrat and Cohen-Tannoudji (2) presented a theoretical analysis of optical pumping which included a discussion of these shifts. They pointed out that the light shifts in the hyperfine transition frequency can be caused by both real and virtual transitions. The frequency shifts due to real transitions are caused by a transfer of coherence from the ground state to the excited state to the ground state the coherence can again be re-transferred back to the ground state. However because the excited state hyperfine splitting is about 1/100th of the ground state hyperfine splitting in ^{85}Rb there is no appreciable ground state frequency shift due to real transitions.

The shifts due to real transitions are of the order of :

$$\Delta\nu_{real} < \frac{\Gamma\gamma}{2\pi(\nu_g - \nu_e)} \qquad (2)$$

where Γ is the pumping rate in Hz

γ is the rate of spontaneous decay from the excited state.

ν_e is the excited state hyperfine splitting.

ν_g is the ground state hyperfine splitting.

Hence, for a pumping rate of 60 Hz

$$\Delta\nu_{real} < 0.3 \text{ Hz}$$

The main contribution of the observed light shift is due to virtual transitions. Such light shifts have been analyzed by Happer et al. (3 & 4), and are shown to be equivalent to a Stark shift caused by the electric field of the pumping light. The electric field of the light polarizes the atomic vapor according to $P \approx \alpha E$ where α is the atomic polarizability and E is the electric field strength of the pumping light. The energy of the interaction of the electric field of the pumping light with the induced polarization is on the order of αE^2. For a typical pumping light flux of 100 μ watts/ cm^2 this interaction causes a frequency shift of the order of 30 Hz. Of course in this order of magnitude approximation the actual spectral profile of the light has been disregarded.

Following Happer (4) a 'light' shift operator for unpolarized light is defined as

$$\delta E = \delta E_0 + h\delta A\, I\cdot J + \delta E_2 \qquad (3)$$

The term δE_o causes an equal displacement in all the ground state hyperfine levels and hence does not cause a net difference in the hyperfine frequency. On the other hand the term $h\delta A\, I\cdot J$ causes frequency shifts for $\Delta F = 1$ transitions and is of interest here. The shift $\delta\nu$ can be written as :

$$\delta\nu = \frac{1}{2}\,\delta A\,(2I+1) = \int_o^\infty \delta_{hfs}(\nu)\varphi\,\nu d\nu \qquad (4)$$

Figures 5 and 6 shown δ_{hfs} vs. frequency displacement for both the D_1 and D_2 resonance lines as a function of temperature for $y = 0$.

The term δE_2 in Eq. (3) is equivalent to a Stark shift caused by the interaction of the atom with the electric field of the light. It is commonly referred to as the "tensor" light shift. In ^{85}Rb the first excited state structure is well resolved and comparable to the Doppler width of the absorption line. It is for this reason that the tensor shift cannot be ignored. Following Happer et al. (3,4) the tensor light shift operator for the field independent transition in ^{85}Rb gives

$$h\delta\nu_t = \langle 3,0|\,\delta E_2\,|3,0\rangle - \langle 2,0|\delta E_2|2,0\rangle \qquad (5)$$

Eq. (10.6) can be rewritten as

$$\delta\nu_t = E_o^2 \int_o^\infty \delta_2(\nu)\,\varphi(\nu)\,d\nu \qquad (6)$$

Figures 7 and 8 show $\delta_2(\nu)$ vs. frequency displacement for both the D_1 and D_2 ^{85}Rb resonance lines for $y = 0$

In order to evaluate $\delta\nu_{hfs}$ and $\delta\nu_t$ from Figures

7 and 8 it is essential to know the spectral profiles and absolute intensities of the pumping lamps. In accordance with the previous assumption of complete filtering of the unwanted lamp component and an additional assumption that the filter cell does not alter the spectral profile significantly it is possible to make realistic approximations of the lamp profile in order to facilitate evaluation of the light shifts. Although the spectral profile of the lamps are known measurements of the absolute spectral intensities is difficult. Instead the total integrated lamp flux was measured with an Eppley thermopile.

Thus

$$\varphi_{obs} = \int_{\nu_1}^{\nu_2} \varphi(\nu)\, d\nu \qquad (7)$$

where ν_1 and ν_2 are the cut off frequencies of the particular interference filter used. Using the known spectral profile, it was assumed that the resonance lines were about 50 mK wide. Hence $\varphi(\nu)$ was approximated by a rectangle 50 mK wide whose height was determined by Eq. (7). Using this approximation, the light shift was evaluated according to Eq. (4) as, a function of light intensity. The results are shown in Figure 9. The contribution of the tensor light shift to the total light shift is about 10% for the D_1 line and about 1.0% for the D_2 line. Figure 9 also shows the measured light shift using both the D_1 and D_2 lines to pump the maser.

Since the ^{85}Rb pumping lamp profile almost overlaps the ^{85}Rb absorption line in the maser, the light shifts predicted by the theory for low light fluxes are small. However if natural rubidium is used as a resonance lamp the lamp profile would be asymmetric and strongly dependent on the filter cell temperature. Hence for a natural rubidium lamp the light shifts would be larger and more sensitive to filter cell temperature than for a ^{85}Rb resonance lamp.

Since the light causes shifts of ≈ 20 Hz for a change of 100% light intensity, the lamp intensity must be stable to 1 part in 10^4 to achieve a maser stability of 1 part in 10^{13}. This lamp stability is obtainable especially in the short term.

* The work was supported wholly by the Joint Services Electronic Programs (U.S. Army, U.S. Navy, and the U.S. Air Force) under contract.

(1) M. Arditi and T.R. Carver, Phys. Rev. 124, 800 (1961).

(2) J.P. Barrat and C. Cohen-Tannoudji, J. de Physique 22, 329 (1961); 22, 443 (1961).

(3) W. Happer and B.S. Mathur, Phys. Rev. 163, 12 (1967).

(4) B.S. Mathur, H. Tang, and W. Happer, Phys. Rev. 171, 11 (1968).

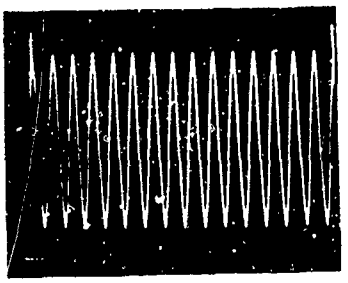


Fig. 1 Cw maser signal as observed at the output of the i.f. strip. The time scale 0.05 μsec/cm.

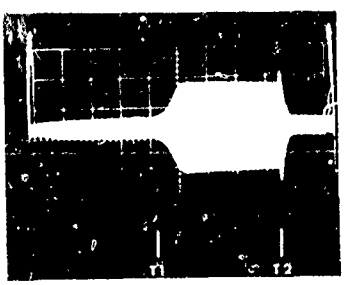


Fig. 2 Transient behavior of maser at the output of the i.f. strip. The pumping light is turned on at t = T1, and is turned off at t = T2. The time scale is 0.1 sec/cm.

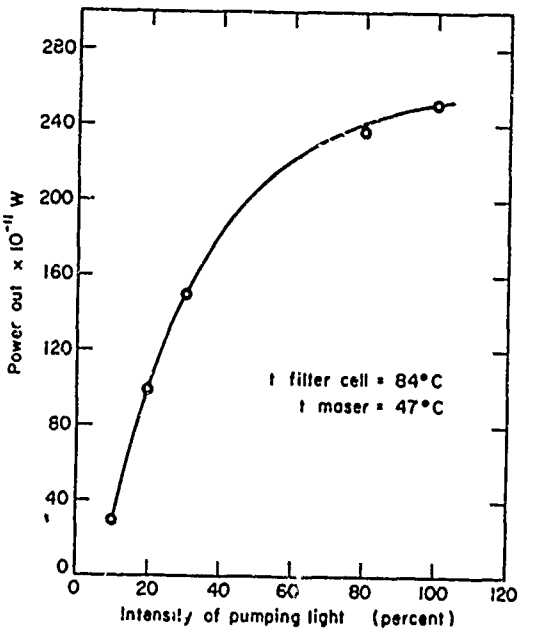


Fig. 4 Power output as a function of pumping light intensity.

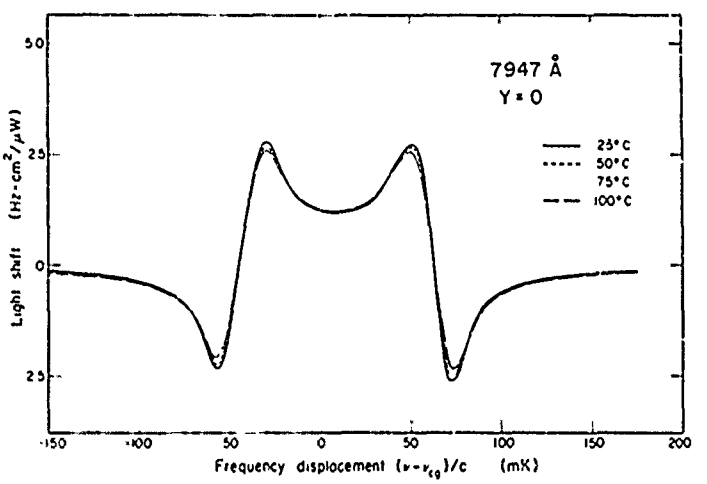


Fig. 5 δ_{hfs} vs. frequency displacement for the D_1 resonance line of ^{85}Rb for y = 0.0.

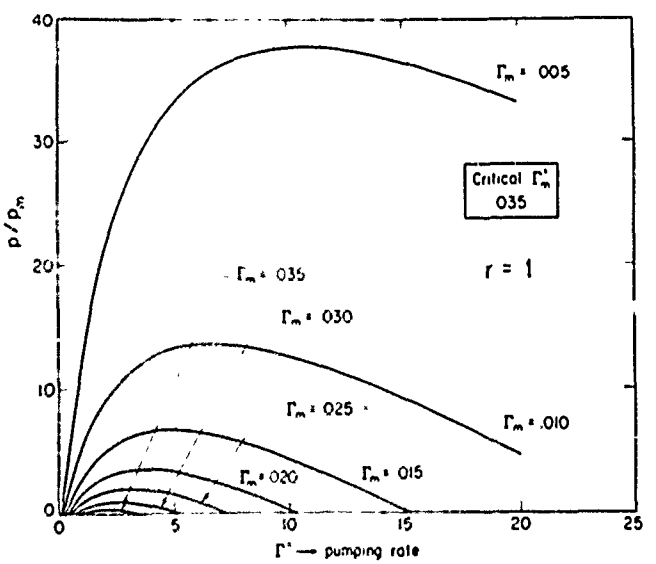


Fig. 3 Normalized power output as a function of the pumping rate for r = 1.

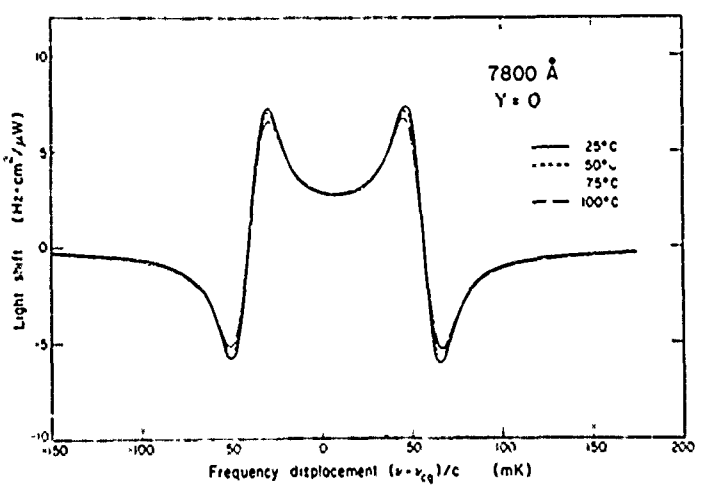


Fig. 6 δ_{hfs} vs. frequency displacement for the D_2 resonance line of ^{85}Rb for y = 0.0.

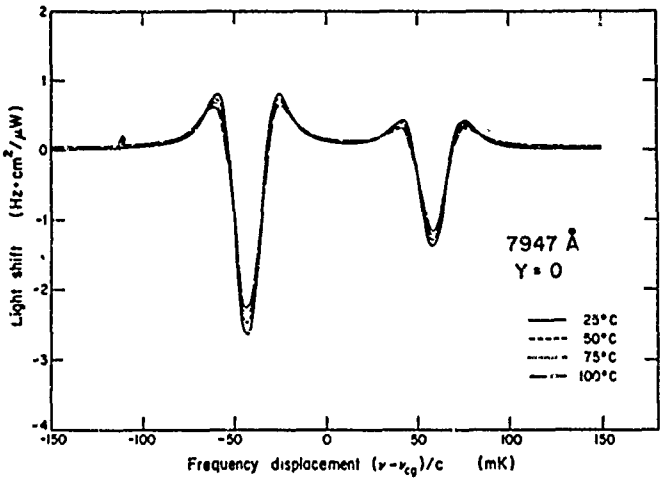


Fig. 7 δ_2 vs. frequency displacement for the D_1 resonance line of ^{85}Rb for y = 0.0.

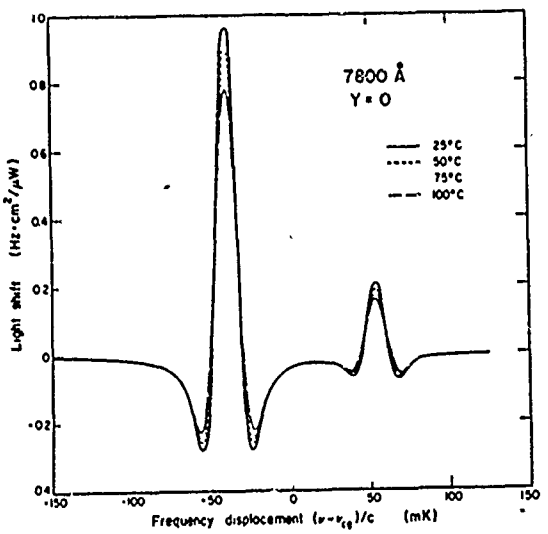


Fig 8. δ_2 vs. frequency displacement for the D_2 resonance line of ^{85}Rb for y = 0.0.

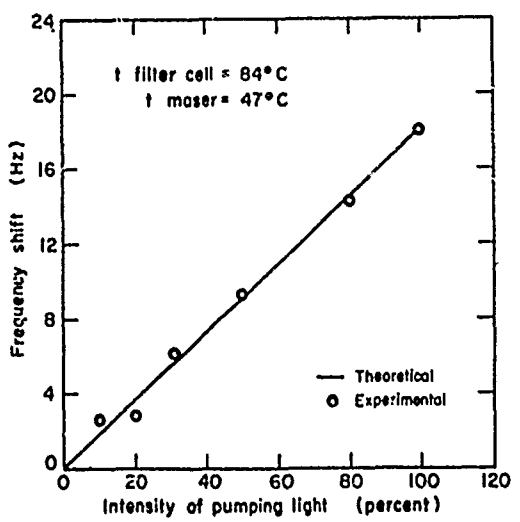


Fig. 9 Comparison of the measured and theoretical light shift; frequency shift vs. relative intensity.

MINIATURISED, RAPID WARM-UP, RUBIDIUM FREQUENCY SOURCE

M. M. Zepler, T. J. Bennett, G. T. Norton and R. E. Hayes

Plessey Company Limited
Roke Manor, Romsey,
Hants, England.

Summary

A prototype small size rubidium frequency source has been produced, which provides a warm up time of less than 5 minutes and overall short and long term frequency stability of better than 1 part in 10^8, which lies between the values achieved in commercial crystal sources and atomic standards.

The new rubidium source uses the same principles as a high grade atomic standard, but the less stringent specification in terms of frequency stability has led to a completely new design approach. The resulting simplification has allowed an experimental model to be built in a volume of just over 100 in^3.

The use of a miniature absorption cell containing natural rubidium enclosed by a single proportionally controlled oven is described. No separate rubidium filter cell is used, and the absorption cell is directly irradiated with microwave power without the need for a waveguide cavity to enclose the cell. The microwave power is obtained by harmonic generation from a crystal oscillator. These features allow the small size and rapid warm-up. Sufficient magnetic shielding is provided by using mu-metal for the equipment case.

Integral output frequencies are provided by a synthesizer, a unique feature of which is the facility of giving small incremental frequency changes to adjust for the tolerance in the manufacture of the absorption cells.

Introduction

A requirement for a high precision frequency source exists in a considerable number of communication systems, both military and civil. This precision is often met by the use of a crystal frequency standard which can provide excellent short-term stability but suffers from long-term drift due to the inevitable ageing of the crystal. If the crystal is continuously maintained at a constant temperature, it is possible to produce a crystal frequency standard which, after an initial running-in period, maintains a stability of better than 1 part in 10^7 over one year. Such a frequency standard is bulky and must, of course, never be switched off.

When it is necessary to achieve a frequency precision better than that which can be provided by a crystal frequency standard, an atomic frequency standard must be employed. Over the past fifteen years several different types have been developed, the most precise of which is capable of giving a long-term stability of better than 1 part in 10^{12}. Even the least accurate and most competitive atomic frequency standard (rubidium vapour) gives a precision of better than 1 part in 10^{10} over a year. It is, however, considerably more bulky and expensive than a precision crystal frequency standard. There are likely to be many applications where a performance intermediate between those provided by available crystal and rubidium frequency standard is adequate.

Investigations were therefore carried out to determine the feasibility of producing a simplified rubidium frequency source to compete with existing crystal frequency standards. Such a source would need to have better long-term stability, a smaller size, and a rapid warm-up time. As a result a prototype model has been built with a long-term stability of better than 1 part in 10^8, in a volume of a little over 100in^3, and with a warm-up time of less than 5 minutes.

Rubidium Frequency Standards

Commercially available rubidium frequency standards use a spectral lamp containing Rb 87, a filter cell containing Rb 85 and an absorption cell containing Rb 87 together with a non-magnetic buffer gas. This gives high efficiency optical pumping and a narrow resonance line[1,2]. Because of the small amount of microwave power available the absorption cell is contained in an H_{01} waveguide cavity tuned to the Rb resonance frequency. The cavity has to have holes or slots to allow the passage of light through it, and its size which is governed by the frequency determines the size of the absorption cell. The thermal capacity of the cavity and two cells sets a lower limit on the warm-up time and places restrictions on the oven design.

To maintain the high order of frequency precision which the standard is capable of providing, very careful attention has to be paid to the three parameters which have an effect on the resonance frequency.

1. Temperature. Double proportionally controlled ovens are required for both the optical assembly and the crystal oscillator to maintain the temperature constant to within about 0.1°C. Because of the temperature dependence of the buffer gas filling, a precise mixture of gases having opposite temperature coefficients is normally used, adding to the manufacturing complexity.

2. Light Intensity. A light intensity monitor and intensity control are desirable to minimize frequency shifts.

3. Magnetic Field. Triple mu-metal shielding is required to keep variations due to external fields within specification.

Simplified Rubidium Frequency Source

A new design approach was adopted in the development of the simplified rubidium source. This was possible for two reasons, first, the reduced specification meant that variations in temperature, magnetic field and light intensity posed less severe problems, and secondly, technology had advanced sufficiently to enable much of the circuitry to be

miniaturised to a considerable extent.

Figure 1 is a block diagram of the simplified rubidium source. It will be seen that the basic principles of a high grade standard are still used. The various parts of the source are now considered in detail.

Optical System

The advances in microwave techniques referred to later have made it possible to dispense with the use of a microwave tuned cavity, so allowing a reduction in the size of the absorption cell, and this, in turn, has led to other simplifications.

Experiments have been carried out using natural rubidium (73 per cent Rb 85, 27 per cent Rb 87) in the absorption cell[3]. We have found that a good resonance signal can be obtained without the need for a separate filter cell. The resonance curve is shown in Figure 2 from which it will be noted that the width of the curve is about 2kHz measured between points of maximum slope. This increased width is in fact a desirable feature in the simplified system as will be seen later; it results from the use of natural rubidium which itself considerably eases the problems in manufacturing the cells.

A large resonance signal is obtained even with the relatively inefficient optical pumping occurring with natural rubidium because a high light intensity is used. This high intensity occurs for three reasons: (a) there is no filter cell; (b) the absorption cell has a small-cross-sectional area and; (c) the light does not have to pass through a perforated cavity wall.

The rubidium spectral source consists of a small glass bulb about 10mm in diameter containing natural rubidium together with krypton as a starter gas. The bulb is supported by a coil which carries an r.f. current at a frequency of around 100MHz.

The light transmitted through the cell is focussed onto a photodiode operating in the reverse-biased mode rather than the more usual photovoltaic mode. It has been found that this gives a considerable improvement in the detected signal-to-noise ratio.

Temperature Control

The cell assembly is maintained at a temperature of approximately 70°C by enclosing it in a single proportionally controlled oven. As mentioned previously it has been the practice to use a mixture of buffer gases to obtain a nearly zero temperature coefficient of resonance frequency. However, it was found that with the high light intensity the temperature coefficient for the nitrogen buffer gas used became negative rather than positive as quoted in the literature.[4] It was therefore inferred that there would be a value of light intensity for which the temperature coefficient was zero. Curves such as those of Figure 3 were plotted from test measurements and it can be seen that there is, for any given light intensity, a temperature range of five degrees centigrade over which frequency stability better than 5 parts in 10^{10} can be attained. A neutral density filter can be included to permit adjustment of the light intensity. As the frequency stability specification has been limited to 1 part in 10^8 a single oven gives sufficient control.

Magnetic Shielding

The magnetic field dependence of the rubidium resonance is given by the expression:

$$f = f_o + \beta H^2 \quad \text{hence} \quad \frac{\delta f}{\delta H} = 2\beta H$$

where f is the resonance frequency in a field of H gauss f_o is the zero field frequency and β is a constant (β = 570 for Rb 87).

It is seen that the change in frequency occurring when the magnetic field varies is proportional to the value of the static field. Thus the value of magnetic field used for Zeeman splitting determines the extent of shielding needed against external fields. In the simplified standard a field of a few milli-gauss is provided by a permanent magnet and a single mu-metal screen is adequate to guard against external field variations of up to ten gauss, without giving a frequency change of greater than a few parts in 10^9. For simplicity, the case containing the rubidium source is made of mu-metal to provide the screen. For test purposes fine tuning can be achieved by rotating the permanent magnet.

Microwave Source

Efficient high order frequency multiplication has become possible with the advent of the step recovery diode, and it is now possible to produce in a single stage of multiplication, sufficient power at 6.834 GHz to observe the rubidium resonance, without the need for the Q magnification of a cavity. In the case the crystal-oscillator frequency has to be an exact sub-harmonic of the rubidium resonance frequency and a method of frequency synthesis involving mixing cannot be used. The use of a Plessey digital synthesizer has resulted in the adoption of a step recovery diode multiplier in a very simple microwave source. The complete microwave source consists of a crystal oscillator at 89.93 MHz, an amplifier providing 500mW and finally a step recovery diode giving multiplication by a factor of 76 to produce about 0.5mW at 6.834 GHz. The rubidium cell is irradiated with the microwave power using a "cheese" reflector.

Automatic Frequency Control (AFC)

The method of obtaining an AFC signal is similar to that used in existing standards. As the resonance curve is now 2kHz wide it has been found that a locking signal can occur when the microwave frequency is 10kHz off resonance. This means that a crystal stability no better than 1 part in 10^6 is required and this is achieved with a crystal oscillator incorporating a simple temperature compensation circuit.

The crystal oscillator is frequency modulated by the signal from a 79Hz oscillator. The output from the photodetector is fed into an amplifier tuned to 79Hz in order to remove the large second harmonic of this frequency which occurs at or near resonance. This tuned amplifier is of the well-known active filter type using integrated circuit operational amplifiers. To avoid problems due to relative frequency drift the 79Hz oscillator uses an identical circuit, with the addition of a feed-back loop. The output of the tuned amplifier is fed to the phase-sensitive detector which uses an output from the 79Hz oscillator as a phase reference. The output from the phase-sensitive detector is used to control the frequency of the crystal oscillator by means of varactor diode tuning. The overall AFC loop gain is about 1000 which gives adequate correction of the open loop crystal frequency.

Frequency Synthesizer

The output from the 89.93 MHz crystal oscillator is stabilized to the rubidium reference, but is at an inconvenient frequency for most applications and is

frequency modulated. A synthesizer is employed to derive standard outputs at 5 MHz, 1 MHz and 100kHz. The basic synthesizer consists of a 5 MHz crystal oscillator, two digital divider chains and a phase comparator, the output from which effectively phase locks the 5 MHz to the rubidium-derived frequency.

Figure 4 is a block diagram of the synthesizer. A conventional phase-locking method is modified by the addition of the divider P, together with the "lose a pulse" (LAP) circuit. This circuit inhibits one pulse at frequency f_R every time it is triggered by an output pulse from the divider P.

Without the LAP circuit:

$$f_o = \frac{N}{M} f_R$$

The expression is modified by the LAP to:

$$f_o = \frac{NPf_R}{MP+1}$$

If suitable values are chosen for the divider ratios M, N, P, changing P by 1 digit enables the frequency f_R to be changed by any desired fractional increment.

If for instance it is desired to change the frequency in steps of 1 part in 10^9 by changing P in steps of 1, then it can be shown that P^2M must be equal to 10^9.

A further consideration of the synthesis shows that the "lose a pulse" occurs at a frequency $\frac{f_R}{PM}$, which appears on the output oscillator as frequency modulation To reduce its effect a filter is included in the phase-lock loop and with a loop cut-off frequency of 1Hz, the "lose a pulse" frequency must not be less than 16Hz to keep the frequency deviation to less than 1 part in 10^8. This gives a further constraint, $PM<7\times10^5$. The filter also effectively removes the 79Hz frequency modulation.

Several compromises had to be made and from a computer calculation the following values were finally selected:

$$M = 643, \quad N = 286, \quad P = 985$$

These values correspond with a rubidium cell having a nominal resonance frequency of 6.834696112 GHz The "lose a pulse" frequency is 17.7 Hz. The increments of adjustment are about 1.6 parts in 10^9.

The prototype model includes a facility for varying the divider ratio, P, between 886 and 1064 by means of two decade switches. This gives an overall adjustment in frequency of approximately 300 parts in 10^7, allowing considerable tolerance in manufacture of the absorption cells.

Power Supply

The whole equipment is capable of operating from a 24V d.c. supply or from a mains power unit about the same size as the frequency source. During warm up 36W is needed, reducing to about 12W during operation.

Performance

Figure 5 is a photograph of the complete prototype rubidium source. The synthesizer (converter) is at present in a separate unit but could of course be packaged with the source itself, increasing the volume by perhaps ten per cent. Further possible refinements are a lock indicator together with a monitor of the loop feedback signal, which gives a measure of frequency deviation.

Measurements have been made under laboratory conditions when it was found that a stability of better than 1 part in 10^8 both short and long term could be maintained over long periods. The warm-up time for this stability is less than 5 minutes as shown in Figure 6, the replica of the output from a chart recorder.

CONCLUSIONS

The prototype model described has not yet undergone any stringent environmental testing, but the design aims of demonstrating the feasibility of a small simplified rubidium frequency source with fast warm-up have been achieved.

ACKNOWLEDGEMENTS

The authors wish to thank all those colleagues who have contributed to the development of the frequency source, in particular to those at the Allen Clark Research Centre, Northants, England, for the production of lamps and absorption cells, and to those at the Frequency and Time Systems Laboratory, West Leigh, Hants, England, for the conception and design of the frequency synthesizer.

The authors are grateful to the directors of the Plessey Company Limited for permission to publish this paper.

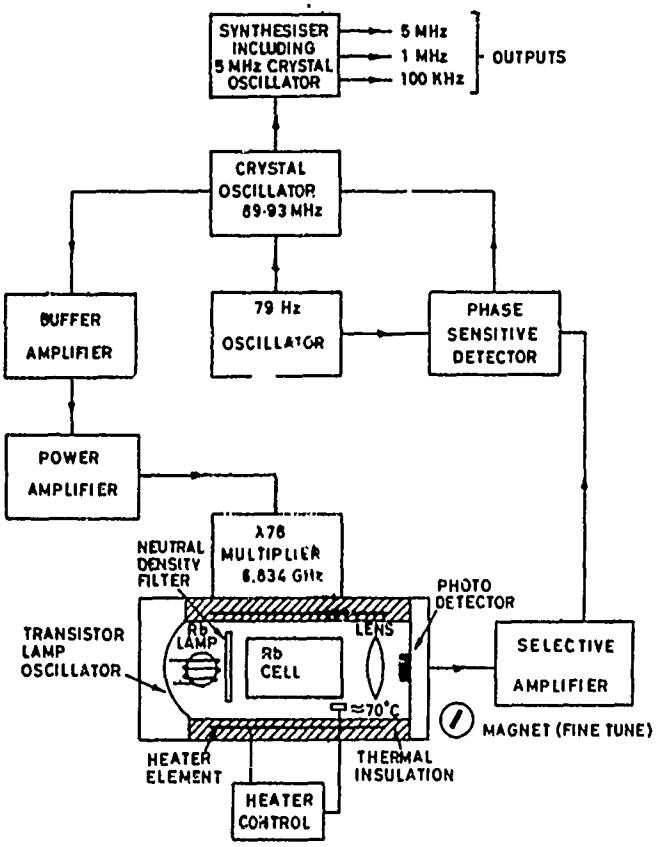
REFERENCES

1. V B Gerard: "Atomic frequency standards using optical pumping of Rb 87 and Cs 133 in gas cells", Brit. J. Applied Phys., August 1962.

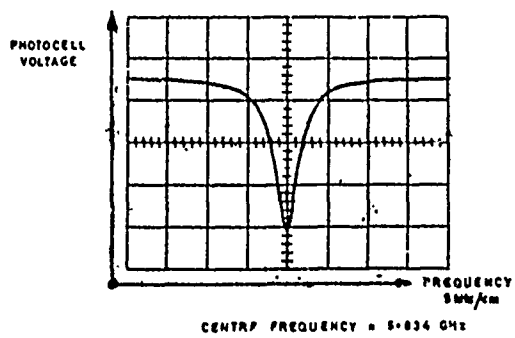
2. P L Bender, E C Beaty and A R Chi: "Optical detection of narrow hyperfine absorption lines", Phys. Rev. Letters, November 1st 1958.

3. V B Gerard: "A gas cell frequency standard using all natural rubidium", PEL Report No.324, DSIR New Zealand, January 1966.

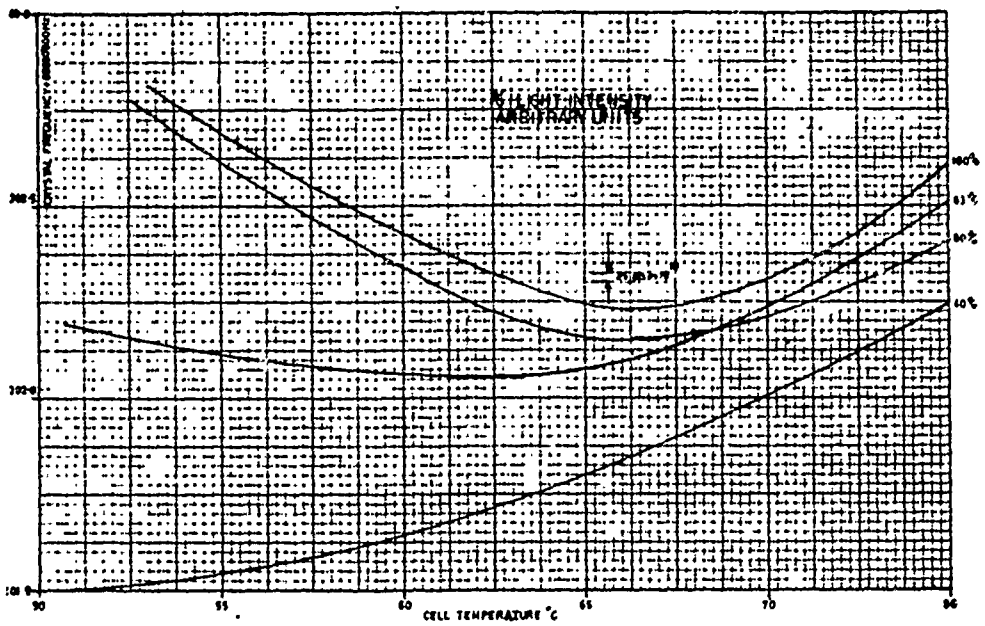
4. M Arditi and T R Carver: "Pressure, light, and temperature shifts in optical detection of 0-0 hyperfine resonance of alkali metals", Phys. Rev., November 1st 1961.



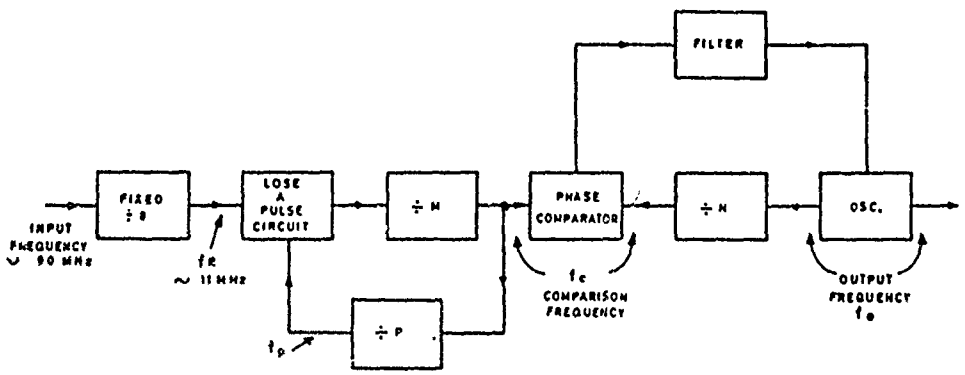
SCHEMATIC OF RUBIDIUM FREQUENCY SOURCE



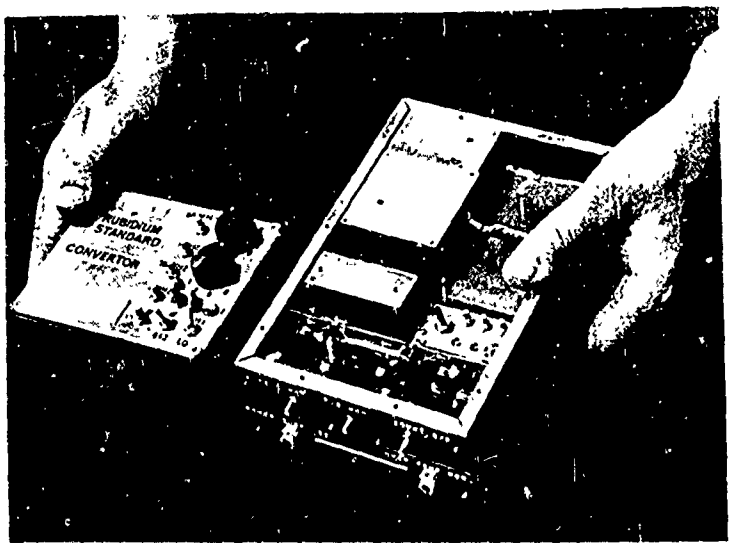
RUBIDUM VAPOUR CELL RESONANCE CURVE.



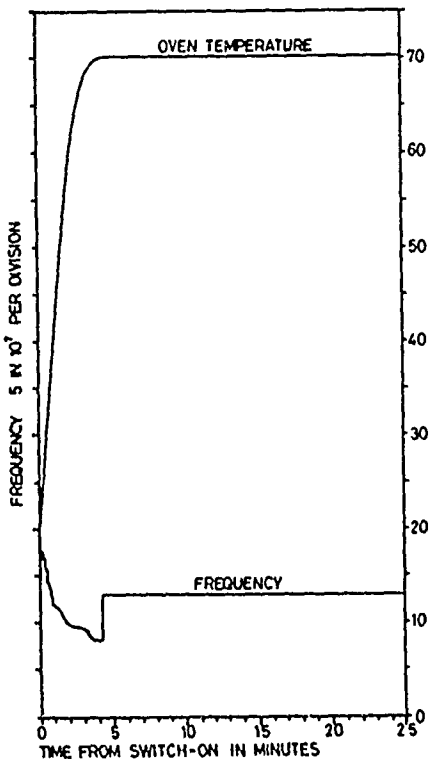
CELL TEMPERATURE CHARACTERISTICS AS A FUNCTION OF LIGHT INTENSITY



FREQUENCY SYNTHESIZER SYSTEM.



EXPERIMENTAL MODEL OF SIMPLIFIED RUBIDIUM FREQUENCY SOURCE WITH SYNTHESIZER



RUBIDIUM SOURCE WARM UP

DISCUSSION OF CAVITY PULLING IN PASSIVE FREQUENCY STANDARDS

J. Viennet, C. Audoin and M. Desaintfuscien
Section d'Orsay du Laboratoire de l'Horloge Atomique
Bâtiment 221 - Faculté des Sciences
91. Orsay - France

Summary

We calculate the cavity pulling factors of passive frequency standards, taking into account the modification of the applied electromagnetic field by the atomic radiation. The calculation is therefore valid for all values of the gain of the atomic medium and gives, in particular, the cavity pulling factor when the oscillation threshold is approached. The effect of saturation of the line by the field is included.

Two expressions of the cavity pulling factor are derived. One applies when the electromagnetic field level in the cavity is measured, and the other when the population of the atomic levels is analyzed. Consequences of these results are discussed.

An experimental verification of the theoretical results is given, when the resonance of the field level in a hydrogen maser cavity is monitored, in conditions wherethe oscillation threshold is approached.

1. Introduction

In frequency standards, the effects of cavity mistunings are well known in two extreme cases.

The first one corresponds, for instance, to the cesium beam tube conditions of operation. There, the modification of the cavity power level by the interaction between the atoms and the high frequency magnetic field is extremely small. The frequency shift of the transition probability is then about $(Q_c/Q_\ell)^2$ times the cavity mistuning [1], where Q_c and Q_ℓ are the quality factors of the cavity and the atomic lines respectively.

Active frequency standards are concerned with the second case where the gain of the atomic (or molecular) medium can be made so high that the atomic radiation sustains the cavity power level. The shift in the oscillation frequency is then Q_c/Q_ℓ times the cavity mistuning [2, 3].

There are now experiments made with passive quantum devices in the intermediate case where the gain of the atomic (or molecular) medium is no longer extremely small. This circumstance can be met in the following examples :

- the ammonia [4], rubidium [5], or hydrogen [6] masers when operated as amplifiers, for spectroscopic measurements.
- the hydrogen storage beam technique proposed by H. Hellwig [7].
- the investigation of saturated absorption of molecular species using an absorbing cell inside the optical cavity [8].

We here consider in more details the influence of cavity mistunings in this intermediate case, giving also the limit of validity of the $(Q_c/Q_\ell)^2$ law.

2. Definition of cavity pulling factors in passive frequency standards

In passive frequency standards, the (angular) frequency ω of the top of the atomic or molecular transition feature is shifted by a cavity mistuning [x]. This offset is characterized by a pulling factor P, defined by :

$$P = (\omega - \omega_o)/(\omega_c - \omega) \qquad (1)$$

where ω_o is the (angular) frequency of the atomic or molecular transition ; $\omega_c - \omega$ is the frequency offset of the cavity, which can be, in practice, approximated by $\omega_c - \omega_o$ in frequency standards.

Two methods of investigating the quantum resonance are currently used, giving rise to different pulling factors.

a) the amplitude of the cavity electromagnetic field is measured while a maser is operated as a spectrometer, or, the intensity of the E. M. field transmitted by the absorbing cell is detected in existing optical frequency standards.
The corresponding pulling factor will then be labelled as P_{Field}

b) The population of one of the two quantum levels involved in the transition is measured by the atomic or molecular beam magnetic resonance methods applied in the cesium beam tube as well as in the hydrogen storage beam tube. In a passive rubidium maser the population difference of the two levels can be monitored by means of transmission of the pumping light through the cell.
The pulling factor will then be labelled as P_{Pop}.

3. The model for the calculation of pulling factors

In atomic and molecular frequency standards, the transition generally occurs inside a cavity, at least in the microwave domain.

[x] other sources of frequency offset are not considered here

3.1 - The atomic medium

We assume that the atomic (or molecular) resonance line in a lorentzian line. This hypothesis is a very good approximation for the Hydrogen and the Rubidium masers [9,5,10]. It leads to simple expressions of pulling factors which can be used as yardsticks to evaluate pulling factors for other shapes of the resonance curves.

The macroscopic behaviour of the atomic medium can then be described by two equations [9]. The first stands for the induced magnetization, and the second one for the saturation of the resonance line :

$$\ddot{M}_1 + \frac{2}{T_2}\dot{M}_1 + \omega_o^2 M_1 = 2 M_3 \omega_o H_1 \qquad (2)$$

$$T_1 \dot{M}_3 + M_3 = -2 T_1 \frac{\overline{H_1 \dot{M}_1}}{\omega_o} + M_o \qquad (3)$$

M_1 is the induced magnetization parameter. It is defined as the value of the magnetic dipole momentum of the whole set of interacting atoms, divided by the Bohr magneton μ_B. M_1 is dimensionless.

M_3 is the total population difference between the two involved quantum levels. M_o is its equilibrium value, in the absence of the high frequency magnetic field perturbation.

H_1 is the driving magnetic field parameter. It is defined as the mean value of the cavity H.F. magnetic field, multiplied by $\mu_B/\hbar$. It is expressed in frequency units.

ω_o is the (angular) transition frequency.

T_1 is the longitudinal relaxation time.

T_2 is the transversal relaxation time. It is related to the Q-factor of the atomic line by :

$$T_2 = 2\frac{Q_\ell}{\omega_o} \qquad (4)$$

3.2 - The microwave cavity

The precession of the induced magnetization excites the useful cavity mode. In passive frequency standards, the cavity is also driven by an external generator in order to interrogate the atomic transition. It can then be shown that the cavity H.F. field is a solution of the following equation [6] :

$$\ddot{H}_1 + \frac{2}{T_c}\dot{H}_1 + \omega_c^2 H_1 = K\ddot{M}_1 - 2\frac{\omega_c}{T_c} H_e \qquad (5)$$

ω_c is the cavity resonance frequency.

T_c is the cavity time constant, related to the cavity Q-Factor by :

$$T_c = 2\frac{Q_c}{\omega_c}$$

H_e is the H.F. field source parameter representing the external generator signal. It gives a cavity response $H_1 = H_e$ when the cavity is free of atoms, and when the excitation frequency ω equals ω_c.

K is a constant defining the coupling between the atomic medium and the cavity H.F. field. It depends on the filling factor, and atomic constants. For the hydrogen maser :

$$K = 4\pi \eta \mu_B^2/\hbar V_c \qquad (7)$$

where η is the filling factor, as defined in the references [3] and [11], μ_B the Bohr magneton and V_c the cavity volume.

3.3 - Useful relations

We introduce the time variations of H_e, H_1, M_1 by :

$$\begin{cases} H_e = p \sin \omega t \\ H_1 = b \cos(\omega t + \varphi) \\ M_1 = m \sin(\omega t + \psi) \end{cases} \qquad (8)$$

For c.w. operation, we get the following useful relations from the atomic-medium equations

- amplitude of the induced magnetic momentum :

$$m = M_o \frac{T_2 b \cos\theta}{1 + T_1 T_2 \cos^2\theta} \qquad (9)$$

- population difference :

$$M_3 = M_o/(1 + T_1 T_2 b^2 \cos^2\theta) \qquad (10)$$

with

$$\begin{cases} \theta = \varphi - \psi \\ \cos^2\theta = \left[1 + T_2^2 (\omega_o - \omega)^2\right]^{-1} \end{cases} \qquad (11)$$

The equation (5) describing the H.F. field leads to two relations which are

$$b = K Q_c m \cos\theta + p \cos\varphi \qquad (12)$$

$$T_c(\omega_c - \omega) b = K Q_c m \sin\theta + p \sin\varphi \qquad (13)$$

They can be transformed, with the expression (9) of the induced magnetic momentum amplitude into :

$$\frac{p}{b}\sin\varphi = T_c(\omega_c - \omega) - K Q_c M_o \frac{T_2 \sin\theta \cos\theta}{1 + T_1 T_2 b^2 \cos^2\theta} \qquad (14)$$

$$\frac{p}{b}\cos\varphi = 1 - K Q_c M_o \frac{T_2 \cos^2\theta}{1 + T_1 T_2 b^2 \cos^2\theta} \qquad (15)$$

4. High frequency field amplitude

The cavity high frequency field amplitude b results from the superposition of the cavity response to the H.F. field source associated with the external generator, and to the atomic radiation.

Eliminating the angle φ in equations (14) and (15), it comes :

$$\left(\frac{p}{b}\right)^2 = 1 + T_c^2(\omega_c - \omega)^2 + \left(\frac{\alpha}{1+S}\right)^2 \cos^2\theta - 2\frac{\alpha}{1+S}\cos^2\theta\left[1 - T_c T_2(\omega_c - \omega)(\omega_o - \omega)\right] \qquad (16)$$

with

$$S = T_1 T_2 b^2 \cos^2\theta \qquad (17)$$

$$\alpha = K Q_c M_o T_2 \qquad (18)$$

The ratio b/p is the gain of the atomic medium defined as the ratio of the actual H.F. amplitude, to the H.F. field amplitude existing as long as the cavity is free of atoms.

$$G = b/p \qquad (19)$$

The gain G can be directly measured by probing the H.F. field amplitudes with a second coupling loop, i.e., by operating the maser as an amplifier by transmission [6].

S is the saturation factor of the line

α is a positive number while the cavity is fed with an excess of atoms in the upper quantum level. It is negative in the inverse case. Its physical meaning can be derived by looking at the gain of the atomic medium in conditions where $\omega = \omega_o = \omega_c$. We get, from equation (16) :

$$G(\omega = \omega_c = \omega_o) = \frac{1+S}{1+S-\alpha} \qquad (20)$$

and

$$G(\omega = \omega_c = \omega_o ; S = 0) = \frac{1}{1-\alpha} \qquad (21)$$

The gain of the unsaturated line becomes infinity for $\alpha = 1$, corresponding to oscillation threshold.

We note that the general expression (16) describes also the power level of the synchronized maser above oscillation threshold [12].

5. Pulling factor for the amplitude of the H.F. field

The derivation of (16) allows to find the excitation angular frequency ω_m giving rise to the extremum value of b (or b^2 as well), the amplitude of the H.F. field in the cavity. For small cavity mistumings we obtain :

$$P_{Field} = \frac{\omega_m - \omega_o}{\omega_c - \omega_m} = Q_c \qquad (22)$$

$$(Q_c + \frac{\alpha}{1+S} Q_\ell)/Q_\ell\left[\frac{\alpha}{1+S}(Q_c + 2\frac{Q_\ell}{1+S}) - \frac{\alpha^2(1-S)}{(1+S)^3} Q_\ell\right]$$

The complete expression is given in the appendix.

Discussion of the result

This result is valid whatever the value of the ratio Q_c/Q_ℓ is. But in frequency standards, we are concerned with the condition $Q_c/Q_\ell \ll 1$, which will be assumed in the following.

Also, the maximum value of S is of the order of unity to avoid excessive line broadening by saturation.

Two asymptotic cases have then to be considered. They are separated by a break corresponding to the following condition :

$$\frac{|\alpha|}{1+S} = \frac{Q_c}{Q_\ell} \qquad (23)$$

When this relation is fulfilled, and $\omega = \omega_o = \omega_c$, the fractional cavity power variation, due to atomic radiation (or absorption) is :

$$\frac{\Delta P}{P} = \left|(\frac{b}{p})^2 - 1\right| \approx 2\frac{Q_c}{Q_\ell} \qquad (24)$$

a) First case : $\frac{|\alpha|}{1+S} \ll \frac{Q_c}{Q_\ell}$

This corresponds to a very low rate of replenishing the population difference in the interaction region. For instance, in the regular hydrogen maser, with $Q_c = 3 \times 10^4$ and $Q_\ell = 10^9$, the atomic flux must be smaller than $3 \times 10^{-5} I_{th}$, I_{th} being the flux at oscillation threshold, for the condition $\alpha < Q_c/Q_\ell$ to be fulfilled.

Indeed, the separation between the two domains take place in conditions far below oscillation threshold..

Equation (22) reduces then to :*

$$P_{field} = (\frac{Q_c}{Q_\ell})^2 \frac{(1+S)^2}{2\alpha}$$

b) Second case : $\frac{|\alpha|}{1+S} \gg \frac{Q_c}{Q_\ell}$

This applies in conditions where a maser is used as a spectrometer, with a convenient value of signal to noise ratio. We get :

$$P_{field} = \frac{Q_c}{Q_\ell} \frac{(1+S)^2}{2 - \alpha + S(2+\alpha)} \qquad (26)$$

For $S \approx 0$:

$$P_{Field}(S \approx 0) = \frac{Q_c}{Q_\ell} \frac{1}{2-\alpha} \qquad (27)$$

At oscillation threshold : $P_{Field}(S \approx 0) = \frac{Q_c}{Q_\ell}$

Figure 1 shows the variation of P_{Field}, as a function of α, for S = 0.

6. Pulling factor for the population

Evaluation of P_{pop} needs as intermediate calculations those performed to obtain P_{Field}. We call ω'_m the excitation frequency for which the difference of population, as well as the population of a given level, reaches its extremum value.

For small cavity mistumings, we get :

$$P_{pop} = \frac{\omega'_m - \omega_o}{\omega_c - \omega'_m} = Q_c(Q_c + \frac{\alpha}{1+S} Q_\ell)/Q_\ell(Q_\ell + \frac{\alpha}{1+S} Q_c) \qquad (28)$$

The complete expression is given in the appendix.

Discussion of the result

As above, we assume that $\frac{Q_c}{Q_\ell}$ is much smaller than unity, and that the saturation factor is not too large. The same asymptotic conditions have to be considered. The break is defined by equation (23) and then, the cavity power variation is given by equation (24).

a) First case : $\frac{|\alpha|}{1+S} \ll \frac{Q_c}{Q_\ell}$

This corresponds to the conditions of operation of a cesium beam tube where the pulling factor is very poor, the atomic flux rather low, and the interaction time short. We get from equation (28) :

$$P_{pop} = (\frac{Q_c}{Q_\ell})^2 \qquad (29)$$

This is the well known cavity pulling factor of the cesium-tube - like frequency standards, which can be more simply derived by assuming that the atomic radiation does not contribute to the cavity power. Our analysis shows that the limit of validity of relation (29) is given by relation (24).

b) Second case : $\frac{|\alpha|}{1+S} \gg \frac{Q_c}{Q_\ell}$

Then, the pulling factor is as follows :

$$P_{pop} = \frac{Q_c}{Q_\ell} \frac{\alpha}{1+S} \qquad (30)$$

Figure 2 shows the asymptotic variation of the pulling factor P_{pop}, as s function of α, for $S \approx 0$. It then varies

* For α vanishingly small, P_{field} is infinity. This means that the cavity beeing free of atoms, $\omega_m = \omega_c$

from $(Q_c/Q_\ell)^2$ to Q_c/Q_ℓ at oscillation threshold, where it reaches the characteristic value of the active frequency standards pulling factor.

Let us compare pulling factors of the oscillating hydrogen maser, and the hydrogen storage beam tube [7].

For the regular hydrogen maser, with $Q_{c1} = 3 \times 10^4$ and $Q_\ell = 10^9$, the pulling factor is $P = 3 \times 10^{-5}$. The atomic flux for oscillation threshold is then about 10^{12} atoms. s^{-1}.

Operating the hydrogen storage beam tube with the same atomic flux, and $Q_{c2} = 10^3$, we have $\alpha = 3 \times 10^{-2} \gg Q_{c2}/Q_\ell$. The pulling factor is then given by equation (30). Assuming S = 0, we get $P_{pop} = 3 \times 10^{-8}$. In this example, the pulling factor is then three orders of magnitude smaller than for the regular hydrogen maser. $[(Q_{c1}/Q_{c2})^2 = 10^3]$ This would only require a rough thermal control of the cavity.

7. Comparison of pulling factors

For given values of Q_c, Q_ℓ and α, assuming S = 0, we can see that, in passive frequency standards

$$P_{Field} > P_{pop} \qquad (31)$$

except for the limiting case $\alpha = 1$ where $P_{Field} = P_{pop}$.

This shows, in this respect, the superiority of methods of resonance frequency measurements based on the observation of level population as being less sensitive to cavity mistuning.

8. Experimental results

We have experimentally verified the expression (26) of P_{Field}, below oscillation threshold of a regular hydrogen maser, for $0,1 < \alpha < 1$. (The parameter α is proportional to the atomic beam flux if T_1 and T_2 are constants). It is then possible to record hyperfine resonance curves with a very good signal to noise ratio, using an I.F. bandwidth of 50 Hz, and a detector time constant of 1 s.

The experimental set-up is shown in figure 3. The maser is used as an amplifier by transmission : one coupling loop is used to fed the external excitation to the cavity, and the other to probe the H.F. magnetic field amplitude in the cavity through.a multiheterodyne receiver followed by a linear detector.

The resonance line is obtained by sweeping the frequency of the xtal oscillator driving the generator.

8. 1 - Measurementof α and Q_ℓ

For a given value of the atomic flux, the values of α and Q_ℓ are directly measured on the recorded line. For S small enough, the value of α is obtained from equation (21) as illustrated in figure 4.

The value of Q_ℓ is related to the linewidth $\Delta\omega$, measured at half height of the line (above the pedestal corresponding to the atom-free cavity), by the following expression which can be derived from equation (16) :

$$T_2^2 (\Delta\omega)^2 = (2 \frac{Q_\ell}{\omega_o})^2 (\Delta\omega)^2 = \frac{(2 - \alpha)^3}{4 - 3\alpha} - 1 \text{ for } S = 0 \qquad (32)$$

T_2 is roughly constant for the experimental values of α. We obtain $T_2 = 0,140$ s. and $Q_\ell = 6,25 \times 10^8$ Q_c equals 48 700.

8. 2 - Pulling factor of the unsaturated line

At first, the validity of the equation (16) has been verified by recording the line, for $\omega_c - \omega_o$, and $(\omega_c - \omega_o)/2\pi = 4.5$ KHz, the condition $S \approx 0$ being experimentally fulfilled. It has been compared with computed values of b/p using the experimentally determined values of α and T_2. Figure 5 shows that the agreement between the calculated and the recorded curves is very good.

Then, for given values of α, the frequency shift of the top of the line has been measured for several values of cavity detuning, at low excitation level ($S \approx 0$). Figure 6 shows the agreement between the measured and the calculated value of P_{Field}.

8.3 - Pulling factor of the saturated line

The saturation factor can be determined from equation (20) by comparing the actual maser gain to the gain obtained for $S \approx 0$ (at very low level of excitation). This determination is illustrated on figure 7.

For a fixed value of α, the pulling factor has been measured for several values of S. Again the experimental points fit the theoretical curve (figure 8).

9. Conclusion

Pulling factors have been calculated for the two ways of measuring an atomic (or molecular) resonance frequency in existing passive frequency standards.

Results apply in all conditions which can be met in these standards, and the validity of known partial results have been quantitatively precised.

Results are valid strictly for lorentzian lines. Work is in progress to cover other possibilities.

10. Appendix

10- 1 - Frequency of the extremum value of the H.F. field amplitude

This frequency is a solution of the following equation :

$$\alpha^2 T_2^2 (\omega_o - \omega) \cos^4\theta \frac{1 - S}{(1+S)^3}$$
$$- \alpha \left[T_c T_2 (\omega_c - \omega) + T_c T_2 (\omega_o - \omega) \right] \cos^2\theta \frac{1}{1+S}$$
$$- 2\alpha T_2^2 (\omega_o - \omega) \cos^4\theta \frac{1}{(1+S)^2} \left[1 - T_c T_2 (\omega_c - \omega)(\omega_o - \omega) \right]$$
$$- T_c^2 (\omega_c - \omega) = 0 \qquad (33)$$

Two solutions may be found for large enough detuning.

10- 2 - Frequency of the extremum value of the population

This frequency is a solution of the equation :

$$T_c (\omega_c - \omega) \left[T_c + \frac{\alpha}{1+S} T_2 \cos^2\theta \right]$$
$$+ T_2 (\omega_o - \omega) \left\{ \frac{\alpha}{1+S} T_c \cos^2\theta + T_2 \left[1 + T_c^2 (\omega_c - \omega)^2 \right] \right\} = 0 \qquad (34)$$

References

1. J.H. Holloway and R.F. Lacey, Congrès International de Chronométrie - Lausanne 1964, Actes p. 319

2. K. Shimoda, T.C. Wang and C.H. Townes, Phys. Rev. 102 (1956) p. 1308

3. D. Kleppner, H.M. Goldenberg and N.F. Ramsey Phys. Rev. 126 (1962) p. 603

4. J.P. Gordon, H.J. Zeiger and C.H. Townes, Phys. Rev. 95 (1954) p. 282

5. F. Hartmann, Journal de Physique, 28 (1967) p. 288

6. C. Audoin, J.L. Duchêne and J.P. Schermann, C.R. Ac. Sci. 268 (1969) p. 1757

7. H. Hellwig, Metrologia 6 (1970) p. 56

8. R.L. Barger and J.L. Hall, Phys. Rev. Letters 22 (1969) p. 4

9. C. Audoin, Revue de Physique Appliquée, 2 (1967) p. 309

10. J. Vanier, Phys. Rev. 168 (1968) p. 129

11. D.Kleppner, H.C. Berg, S.B. Crampton, N.F. Ramsey, R.F.C. Vessot, H.E. Peters and J. Vanier, Phys. Rev. 138 (1965) p. A. 972 - See Note n° 27

12. C. Audoin and J. Viennet, Colloque International de Chronométrie - Paris 1969 - Actes p. A5

Figure captions

1. Asymptotic variation of P_{Field}, as a function of α, for S = 0 (logarithmic scales)
2. Asymptotic variation of P_{pop}, as a function of α, for S = 0 (logarithmic scales)
3. Experimental Set-up
4. Measurement of α - Definition of Δω
5. Comparison between recorded and computed lines
6. Pulling factor, vs α. The full line is the theoretical curve
7. Measurement of S
8. Pulling factor vs saturation factor. The full line is the theoretical curve

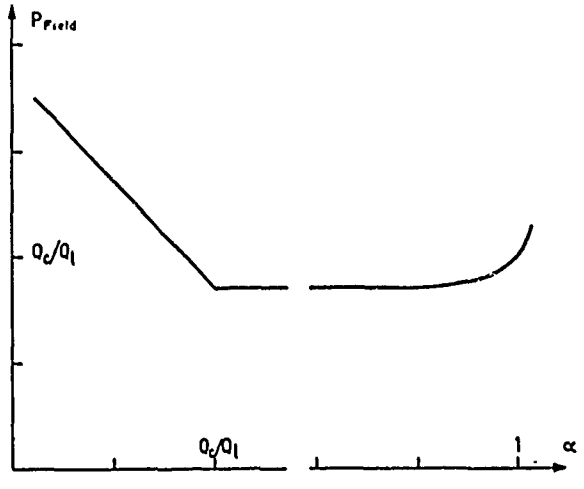


Fig.1_Asymptotic variation of P_{Field} for S=0 [Logarithmic scales]

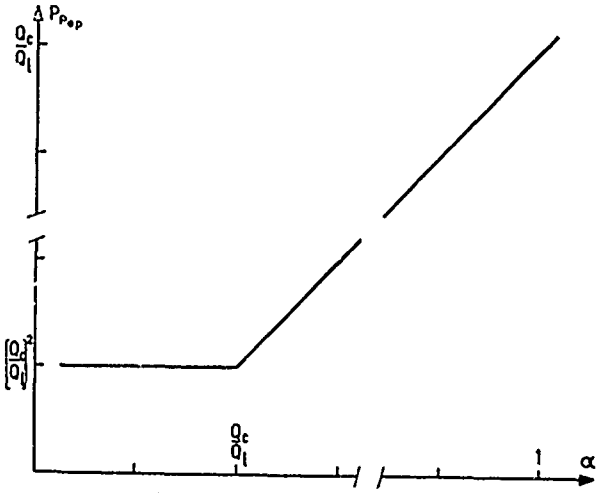


Fig.2_Asymptotic variation of P_{pop} for S=0 [Logarithmic scales]

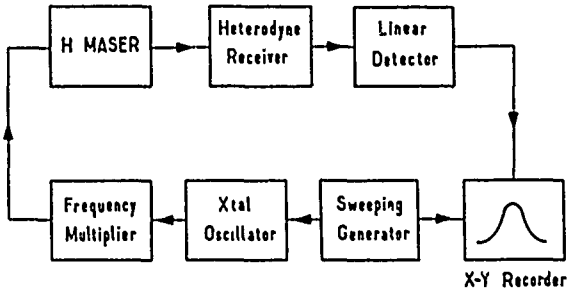


Fig.3 _ Experimental Set-up

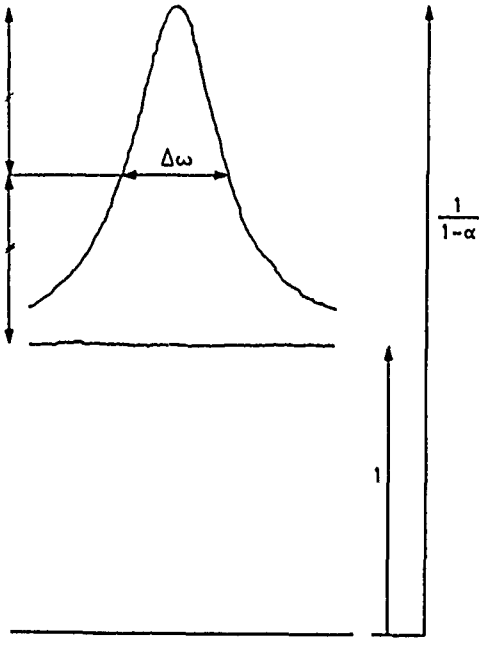


Fig.4 _ Measurement of α _ Definition of Δω

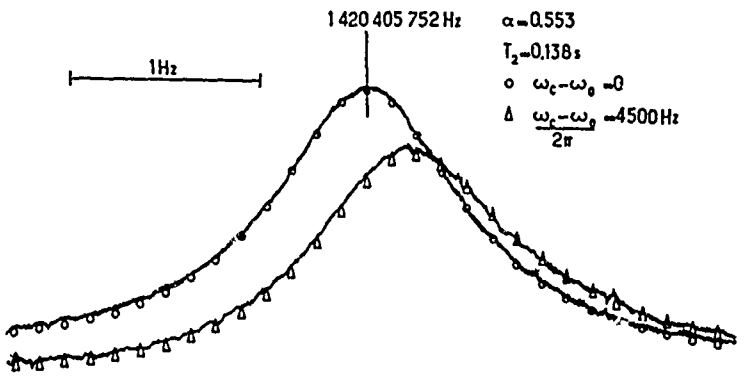


Fig.5 — Comparison between recorded and computed lines

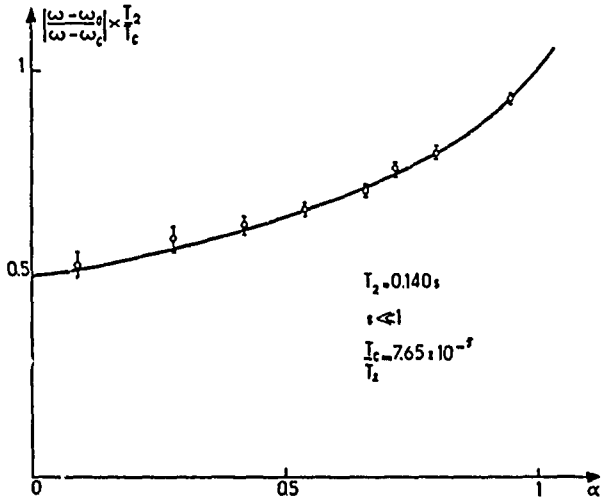


Fig.6 — Pulling factor vs α

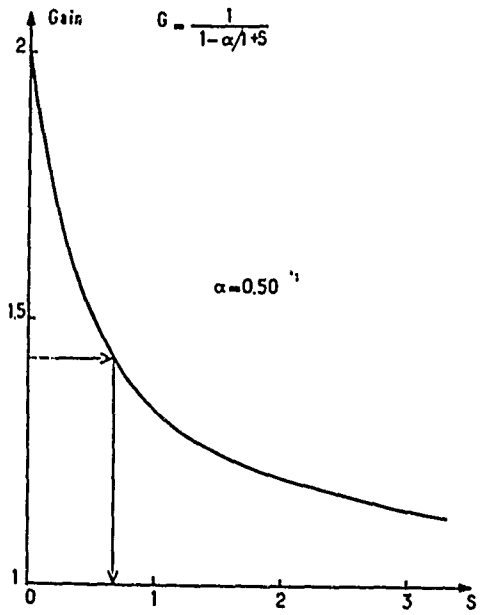


Fig.7 — Measurement of S

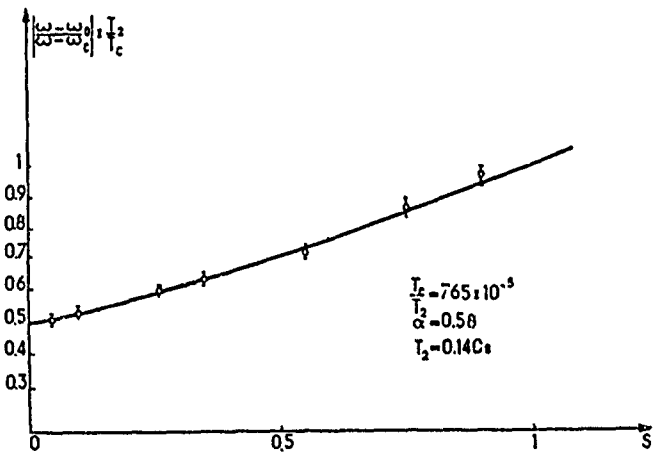


Fig.8 — Pulling factor vs saturation factor

HYDROGEN MASER WALL SHIFT EXPERIMENTS AT THE NATIONAL RESEARCH COUNCIL OF CANADA

D. Morris
Physics Division, National Research Council,
Ottawa, Canada

Summary

One of the limiting factors which restrict the hydrogen maser to the role of a secondary frequency standard is the uncertainty in the experimental determination of the wall shift. This paper describes a series of experiments carried out at the National Research Council of Canada, using two hydrogen masers, to investigate the wall shift of five spherical storage bulbs of different diameters, each coated by the same technique with FEP Teflon from the same batch. The theoretically predicted inverse relationship between wall shift and bulb diameter was found to hold experimentally with very small point scatter for all five bulbs. Results showing the dependence of wall shift on bulb diameter and the variation of wall shift with temperature from 23°C to 35°C are given. Measurements made on the largest storage bulb at the beginning and end of the series of experiments allow some estimate to be made of the stability of the wall shift over a period of several months. Measurements indicating a change in wall shift of not more than 2 parts in 10^{13} over a period of seven months for a storage bulb transported between Japan and Canada are also discussed.

Comparison of the frequency of one of the masers with that of the NRC primary cesium beam frequency standard (Cs III) has given a value of the unperturbed hydrogen hyperfine transition frequency of 1 420 405 751.770 ± 0.003 Hz, which is in satisfactory agreement with the most recent determinations by other workers.

Introduction

In the last few years much effort has been expended in various laboratories on the accurate evaluation of the frequency shift arising from the effect of the confining walls of the storage bulb in the hydrogen maser. Despite this work, uncertainties in the determination of the wall shift still limit the hydrogen maser to the role of a secondary frequency standard. The reduction or elimination of this uncertainty is necessary before the maser can be used as a primary standard of frequency. A number of techniques have been proposed and tested for reducing this uncertainty. These include the use of a very large storage bulb with correspondingly reduced wall shift,[1] the use of a deformable bulb,[2] and the use of a bulb coated with PTFE and operated at a temperature of about 80°C at which the wall shift is zero.[3] However, the method of determining wall shift by using a number of storage bulbs of different diameters, coated by the same technique with material from the same batch, still appears to offer certain advantages. This method obviates difficulties concerned with the large physical dimensions and the auxiliary amplifiers necessary with the large storage bulb maser, the mechanical instabilities associated with the deformable bulb, and problems associated with the relatively high temperature operation of the zero wall shift maser. Also, despite the extended extrapolation that is required this method can still constitute an accurate method of determining the wall shift correction. Recently, Essen[4] has described a method using bulbs with two different coating materials to reduce the extrapolation error further.

Measurement of wall shift

At the National Research Council of Canada two hydrogen masers have been in operation since 1967.[5] These have been used as stable reference oscillators in the evaluation of the NRC long beam primary cesium standard Cs III and also have been examined as possible primary frequency standards. This paper describes a series of measurements of the wall shift of five storage bulbs of different diameters, each of which was coated by the same technique. Both masers were employed, one of them (H2) being used as a stable reference. This maser employed a storage bulb 16.1 cm in diameter coated with FEP Teflon, and with a removable 5 mm diameter solid Teflon plug in the top of the bulb. The Teflon plug had been employed to aid in alignment of the maser. The other maser, H1, which is essentially identical in construction to H2 and originally used a storage bulb of the same type, was fitted with a series of 5 spherical storage bulbs, all coated with the same batch of FEP 120 Teflon dispersion and using the same coating and curing techniques for all the bulbs. The bulbs ranged in diameter from 9.8 to 16.2 cm and in each bulb the entrance hole consisted of a simple aperture approximately 2 mm in diameter. No solid Teflon parts were used in these bulbs.

Each bulb was chemically cleaned with nitric acid and then coated with a diluted dispersion of Du Pont FEP 120 Teflon. Since differences in properties may occur from one batch to another, it should be noted that this material was purchased in 1965. The dispersion was allowed to dry on the surface of the bulb which was warmed externally while dry nitrogen was passed in through the entrance hole. The bulb was then put in a furnace and heated to 360°C and maintained at this temperature for 20 minutes with a continuous slow flow of air entering the bulb. It was found necessary to use either three or four layers of Teflon in each bulb in order to obtain complete coverage of the surface without pinholes. Each bulb was installed in maser H1 a few days after it had been coated.

The cavities of both the masers were temperature controlled at 28.8°C. In the experiments, maser H1 was tuned by means of spin-exchange linewidth broadening, using H2 as the stable reference, and then H2 was tuned using H1 as reference. From the beat frequency between the masers, after correction for magnetic field offsets, second-order Doppler shifts and the small cavity tuning corrections determined for each maser, the difference in the wall shifts ($\delta f_{w1} - \delta f_{w2}$) was determined. Here δf_{w2} is the fixed wall shift of the bulb in maser H2 and δf_{w1} refers to the wall shift of the bulbs used in maser H1. For each storage bulb used in H1 seven or eight in-

Table I

Relative wall shift of bulbs at 28.8°C

Bulb	Diameter (cm)	Inverse diameter (cm^{-1})	No. of Teflon layers	Date of measurement	No. of tunings	Relative wall shift $\delta f_{w1} - \delta f_{w2}$ (mHz)
B	16.27	0.0615	4	April-May 1970	7	+ 6.06 ± 0.25
				Feb. 1971	7	+ 6.03 ± 0.07
C	14.02	0.0713	3	June-July 1970	8	+ 2.45 ± 0.25
D	12.59	0.0794	3	July 1970	7	- 1.53 ± 0.09
E	10.84	0.0923	4	Sept. 1970	8	- 7.62 ± 0.38
F	9.78	0.1022	4	Nov.-Dec. 1970	8	- 11.52 ± 0.23

dependent tunings of the masers were carried out to obtain $(\delta f_{w1} - \delta f_{w2})$. The results are shown in table I, together with the standard deviations from the mean.

The approximate formula[6] for the wall shift is

$$\delta f_{w1} = \frac{K}{D}\left[1 + a(t - t_1)\right]$$

where D is the bulb diameter, t is the bulb temperature and the parameters K and a are defined at the reference temperature t_1. Therefore, if values of $(\delta f_{w1} - \delta f_{w2})$, determined at a constant temperature, are plotted against $\frac{1}{D}$, both K and δf_{w2} can be determined. Experimentally, as shown in fig. 1, the dependence of the wall shift difference on inverse bulb diameter is linear with all five points lying close to the straight line. A weighted least-squares fit to the data gave K = -436 ± 14 mHz cm at 28.8°C, and δf_{w2} = -33.0 ± 1.2 mHz.

To compare this value of K with values found by other workers at different reference temperatures, it was necessary to determine the temperature coefficient of the wall shift. The relative wall shift of bulb B was determined at 28.8°C and 34.8°C, and that for bulb F was determined at 23.0°C, 28.8°C and 34.8°C, the latter being the upper temperature limit attainable with the present oven design. From the data the temperature coefficient, a, was found to be $(-9 \pm 1) \times 10^{-3}°C^{-1}$ at 28.8°C.

If $K(t_1)$ and $a(t_1)$ are the values of K and a determined at a reference temperature t_1, and (t_2) and $a(t_2)$ are those determined at another reference temperature t_2, then it can be shown from the wall shift formula given above that

$$K(t_2) = K(t_1)\left[1 + a(t_1)\cdot(t_2 - t\right]$$

and $$a(t_2) = \frac{K(t_1)\cdot a(t_1)}{K(t_2)}$$

A reference temperature of 40°C is often used as a basis for comparing wall shift results. Therefore, from the present measurements, if t_1 is taken as 28.8°C and t_2 as 40°C, it is found that K = 390 ± 18 mHz cm and $a = (-10 \pm 1) \times 10^{-3}°C^{-1}$ at 40°C. These are compared with other values given in the literature for FEP Teflon at 40°C in table II. The results from the present work are in closer agreement with those of Zitzewitz et al than with those of several other workers. It seems that the values obtained are dependent on the actual batch of Teflon dispersion used.

The presence of the solid Teflon plug in the top of the storage bulb used on maser H2 resulted in an increase in wall shift compared to that of a similar diameter bulb without any solid Teflon parts. This is shown in fig. 1 in which a value is shown for bulb A in maser H1. This bulb was of similar diameter to that in H2 and also employed a Teflon plug. This bulb gave a value of relative wall shift, $(\delta f_{w1} - \delta f_{w2})$, of + 1.18 ± 0.26 mHz at 28.8°C. As seen from fig. 1, the increase in wall shift due to the presence of the Teflon plug was 4.9 mHz or 3.4 parts in 10^{12}.

Determination of the hydrogen hyperfine transition frequency

During part of the time that the wall shift exper-

Table II

Comparison of wall-shift results found for FEP Teflon

Author	Year of measurement	No. of bulbs	K (mHz cm) at 40°C	a ($°C^{-1}$) at 40°C
Vanier & Vessot[6]	1963-1964	6	-528 ± 5	$(-5 \pm 1) \times 10^{-3}$
Mathur et al[7]	1964	4	-500 ± 77	Not measured
Zitzewitz et al[8]	1968-1969	18	-347 ± 7	$(-12 \pm 1) \times 10^{-3}$
Menoud & Racine[9]	1968-1969	2	-515 ± 28	Not measured
Morris	1970-1971	5	-390 ± 18	$(-10 \pm 1) \times 10^{-3}$

Table III

Comparison of recently published values of the hydrogen hyperfine frequency

Author	Year of measurement	Laboratory	Cesium Reference	Hydrogen frequency (Hz) 1 420 405 751 +
Menoud & Racine[9]	1969	LSRH	Commercial unit	0.778 ± 0.004
Hellwig et al[10] (experiment 1)	1970	Harvard University and NBS	NBS lab. standard	0.769 ± 0.002
Hellwig et al[10] (experiment 2)	1970	NBS	NBS lab. standard	0.767 ± 0.002
Essen et al[4]	1970	NPL	NPL lab. standard & commercial units	0.767 ± 0.001
Morris (this paper)	1970-71	NRC	NRC lab. standard	0.770 ± 0.003

iments were being carried out, the frequency of the reference maser H2 was measured with respect to the NRC 2.1 m cesium standard Cs III. The frequency comparison circuitry is shown in fig. 2. The signal supplying the frequency multipliers is derived from a 5 MHz crystal oscillator, which may be either free-running or may be locked to the cesium resonance. In the latter case, the use of the synthesizer and multiplier shown allow the frequency of either maser to be compared with that of Cs III. Therefore, it was possible to compare the frequencies of H2 and Cs III at the same time as the frequencies of H1 and H2 were compared. Seven frequency comparisons between H2 and Cs III were carried out, each on different days, during September and October 1970 after the tuning correction for H2 had been determined in each case. From the results of these frequency comparisons and the value of the wall shift of maser H2, δf_{w2}, obtained from the extrapolation in figure 1, the value of the unperturbed hydrogen hyperfine transition frequency was found to be 1 420 405 751.770 Hz. This value is based on corrections at present applied to both H2 and Cs III. The error limit for Cs III is $\pm 1.5 \times 10^{-12}$, based on the most recent re-evaluation[11] and including the new value of second-order Doppler shift;[12] that for H2 is about $\pm 1.0 \times 10^{-12}$. There is an additional error of $\pm 0.5 \times 10^{-12}$ resulting from random errors in the frequency comparison. The value for the hydrogen frequency has therefore a possible error limit of about $\pm 1.9 \times 10^{-12}$ or ± 0.003 Hz. In table III recently published values of the hydrogen frequency are given. The only values shown in the table are those which were obtained by an independent determination of wall shift at the same time as the hydrogen-cesium frequency comparison was carried out. With the exception of the LSRH result, the agreement between the present NRC determination and the most recent measurements by other laboratories is satisfactory.

Stability of wall shift with time

Exposure of a Teflon-coated storage bulb to air for extended periods might cause a change in wall shift due to chemical changes at the surface. Some estimate of the magnitude of such an effect can be obtained from the data shown for bulb B in table I. This bulb was first installed in maser H1 in April 1970 and removed in June 1970. In April and May the measured difference in wall shift between this bulb and that in the reference maser was + 6.06 ± 0.25 mHz. The bulb was then stored in air until February 1971, when the measurements on the other four bulbs had been completed, and was then re-installed in H1 and the difference in wall shifts was measured again. The value obtained was + 6.03 ± 0.07 mHz. During this period, the reference maser was kept under vacuum. Therefore, within the combined error limits of the measurements, ± 0.26 mHz ($\pm 1.8 \times 10^{-13}$), it is reasonable to assume that exposure to the air for eight months did not cause a change in the wall shift of bulb B.

Additional evidence for the stability of the wall shift of FEP Teflon coatings arises from an experiment carried out jointly by the National Research Council and by the Radio Research Laboratories, Japan. In this experiment, a storage bulb (RRL #453), 15 cm in diameter, was coated with FEP Teflon by RRL and installed in their maser RRL-H2. A measurement of the frequency of this maser was made with respect to the RRL commercial cesium

Table IV

Measurements of maser frequency with RRL bulb #453 at 23°C

Date	Laboratory	Maser frequency (Hz) (not corrected for wall shift)
Aug. 1970	RRL	1 420 405 751.7376 ± 0.002
Oct. 1970	NRC	1 420 405 751.7374 ± 0.0023
Feb. 1971	RRL	1 420 405 751.7373 ± 0.002

standards in August 1970. The storage bulb was brought to NRC by Dr. Y. Saburi and installed in maser NRC-H1 in October 1970. Measurements of the maser frequency with respect to the cesium standard Cs III were made, using maser H2 as transfer standard. The storage bulb was removed from maser H1 in November 1970 and returned to Japan by air. It was re-installed in RRL maser H2 and the frequency again checked. All measurements were made at a storage bulb temperature of 23°C. The results are shown in table IV. The values given are the maser frequencies corrected for magnetic field offsets, tuning corrections and second-order Doppler shifts, but not for wall shift. The close agreement between the three measurements indicates both the lack of any significant change (i.e., not more than 2×10^{-13}) in wall shift over a period of seven months and the close agreement between the cesium reference standards maintained by NRC and RRL.

Conclusions

It has been shown that the use of a number of bulbs of different sizes, coated with the same batch of material using the same technique, is a valid method for the measurement of wall shift, although it requires lengthy and tedious experimental effort. Despite the lack of agreement in the results obtained for the wall shift parameter K for FEP Teflon by different workers, determinations of the hydrogen transition frequency carried out at the same time as wall shift determinations involving several storage bulbs have shown good agreement between different laboratories.

Repeated measurements have shown that the change in wall shift of a particular storage bulb is probably not more than a few parts in 10^{13} per year.

Acknowledgements

I wish to thank Dr. A.G. Mungall, Mr. H. Daams and Mr. R. Bailey for their assistance and advice during the course of this work. I also wish to thank Dr. Y. Saburi of RRL, Japan, for the opportunity to participate in the co-operative wall shift experiment, and for his permission to report the results in this paper.

References

1. Uzgiris, E.E. and Ramsey, N.F., Phys. Rev. A. 1, 429 (1970).

2. Debély, P.E., Rev. Sci. Instr. 41, 1290 (1970).

3. Vessot, R.F.C. and Levine, M.W., Metrologia 6, 116 (1970).

4. Essen, L., Donaldson, R.W., Bangham, M.J., and Hope, E.G., Nature, 229, 110 (1971).

5. Mungall, A.G., Morris, D., Daams, H. and Bailey, R., Metrologia 4, 87 (1968).

6. Vanier, J. and Vessot, R.F.C., Metrologia 6, 52 (1970).

7. Mathur, B.S., Crampton, S.B., Kleppner, D. and Ramsey, N.F., Phys. Rev., 158, 14 (1967).

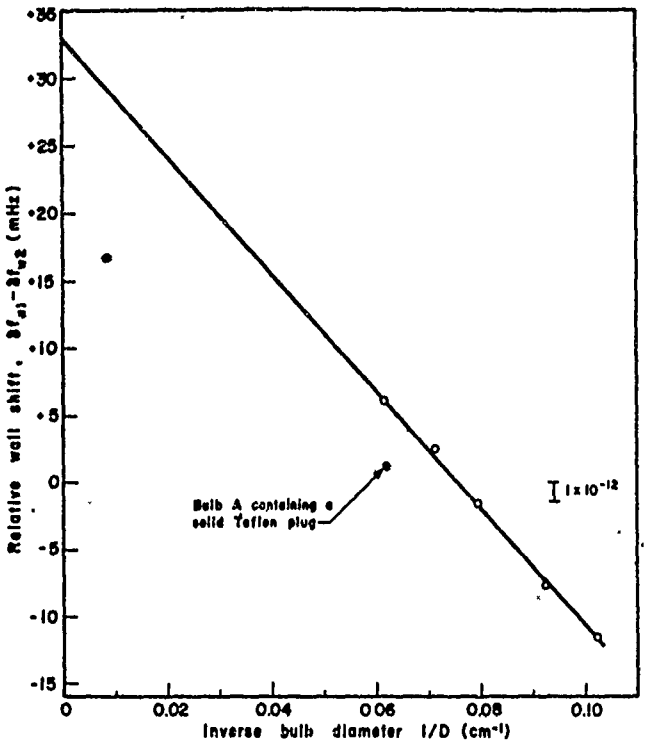
8. Zitzewitz, P.W., Uzgiris, E.E., and Ramsey, N.F., Rev. Sci. Instr., 41, 81 (1970).

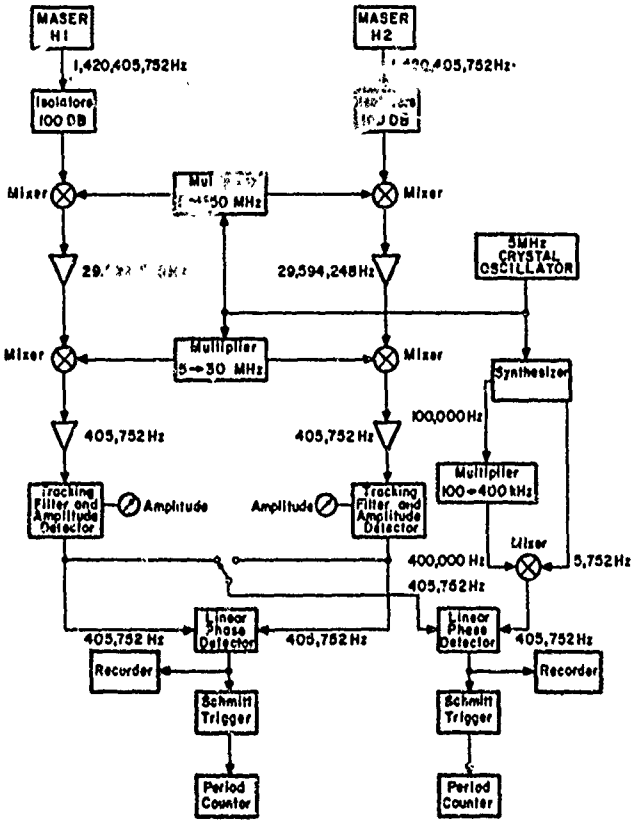
9. Menoud, C. and Racine, J., Z. angew. Math. Phys. 20, 578 (1969).

10. Hellwig, H., Vessot, R.F.C., Levine, M.W., Zitzewitz, P.W., Allan, D.W. and Glaze, D.J., IEEE Trans. Instr. and Meas., IM-19, 200 (1970).

11. Mungall, A.G., Bailey, R., Daams, H. and Morris, D., Metrologia 4, 165 (1968).

12. Mungall, A.G., to be published in Metrologia 7, April 1971.


Relative wall shift, $\delta f_{s1} - \delta f_{s2}$ (mHz)
+.35
+.30
+.25
+.20
+.15
+.10
+.5
0
-.5
-.10
-.15
0
0.02
0.04
0.06
0.08
0.10
Inverse bulb diameter 1/D (cm⁻¹)
Bulb A containing a solid Teflon plug
I 1 x 10⁻¹²


MASER H1
MASER H2
1,420,405,752Hz
Isolators 100 DB
Mixer
Multiplier 5→30 MHz
29,594,248Hz
5MHz CRYSTAL OSCILLATOR
Synthesizer
100,000Hz
405,752Hz
Multiplier 100→400 kHz
Tracking Filter and Amplitude Detector
Amplitude
400,000 Hz
5,752Hz
Linear Phase Detector
Recorder
Schmitt Trigger
Period Counter

HYDROGEN MASER FREQUENCY STANDARD

C. Finnie
R. Sydnor
A. Sward

Jet Propulsion Laboratory
Pasadena, California 91103

Summary

Two hydrogen maser frequency standards were designed and fabricated at the Jet Propulsion Laboratory. They have been in field operation for approximately 1½ years. This paper describes maser design and construction details which were optimized to meet deep space tracking, data acquisition, reliability and maintenance requirements.

Good short term stability (1 part in 10^{-13} @ 5 seconds) was achieved by the use of a high hydrogen flux level, with a resulting output power of approximately -85 dbm. The frequency stability for averaging times of 50 seconds and higher is 1 part in 10^{-14}. Hydrogen pumping speed has been improved using special hydrogen pump elements developed for the Atomic Energy Commission.

In order to synthesize the required output frequencies for the Deep Space Net, a phase-locked loop receiver is utilized. Among many special features incorporated in the receiver is a new type of frequency synthesizer with 10^{-8}Hz increments allowing the masers to be offset in frequency as little as 7 part in 10^{-18}. Special techniques were used in designing the RF modules in order to minimize PM noise, AM noise, and long term phase drift.

Frequency stability measurements have indicated very close correlation between the predicted short term stability and its measured value, as well as excellent long term stability. Some of the other things to be discussed in the paper are:

1. Thick wall Cer-vit cavity design
2. Thick evaporated copper cavity coating
3. Features which make field maintenance possible with minimum equipment and personnel
4. A proportional digital cavity tuner using medium scale integrated circuitry
5. Microwave cavity reflectometer system
6. Optimum noise bandwidth of receiver
7. Techniques used to minimize PM and AM noise in RF modules
8. Low phase noise measurement techniques
9. Long and short term frequency measurements of hydrogen masers and receivers
10. Correlation between predicted and measured stability
11. Maser reliability

Design Features of Maser Physics Unit

The JPL Hydrogen Masers were designed to meet a goal of 1×10^{-13} frequency stability from 5 sec to 10^6 seconds. In addition it was desired to achieve a short term stability as good as practicable to obtain narrow RF linewidth. The units were to be field serviced by means of replaceable modules insofar as was possible. Since the units were to operate in a field environment as opposed to a laboratory environment, consideration of magnetic and thermal transients was necessary. These various requirements dictated to a large degree the overall design and configuration of the unit.

The maser physics unit has been physically separated from the control and receiver electronics in the JPL maser. This allows the physics unit and electronics rack to be interchangeable by a simple cable interface. Also, electronics maintenance and repair are physically removed from the mechanical shock and magnetic field sensitive maser physics package. The separation simplifies the removal of various components of the physics unit for maintenance. The design goals for the physics unit were to optimize short term stability for telemetry applications and to develop a reliable unit which can be easily maintained in the field. Hydrogen maser short term stability (less than 100 sec) is limited by the ratio of the signal power delivered to the maser receiver and the noise contributed by the receiver, cavity and cavity isolating network. Increasing the maser output power is the only practical means of making significant improvements in short term stability.

Two conditions, in particular, limit the output power of a practical maser: the first is a gain saturatic produced by spin exchange collision broadening the hydrogen resonant linewidth. The second is the ion pump life time determined by its capacity to store pumped hydrogen. The JPL masers make use of special ion pump elements developed for the A.E.C. by the Ultek Corporation. This is shown in figure 1. These elements use 3 mm. thick titanium plates on both sides of the pump anodes. The elements are supported on flexible insulator mounts which make the unit less vulnerable to failure from plate warpage. Two sets of these elements have operated in JPL masers for greater than one year without appreciable plate distortion. These pumps were operated at approximately 3×10^{-6} torr and have a 200 liter/sec pumping speed for eight elements. This is equivalent to a hydrogen flux of 1.8×10^{17} molecules/sec., producing a maser output power of -85 dbm. The ion pump current is 4 ma. at 4300 V. Standard ion pumps at this flux level have a hydrogen lifetime less than 6 months - limited by plate warpage and insulator breakage. Line broadening (2-3 Hz) resulting from high maser output power requires exceptional cavity frequency stability. To achieve this, the cavity cylinder is fabricated from Cer-vit material which has a temperature coefficient of approximately 10^{-7}. The cavity is surrounded by an oven within the vacuum chamber. All microwave components in the receiver input circuitry are also temperature controlled. The cavity oven is controlled to $46^{\circ}C \pm .001^{\circ}C$; the microwave components to $46^{\circ}C \pm .010^{\circ}C$, and the maser vacuum chamber and ion pump to $45^{\circ}C \pm .1^{\circ}C$. The Cer-vit cavity is copper coated using a thick evaporation technique developed by Hubert Erpenbach at JPL. The cavity has an unloaded "Q" of 57,000 and is operated at a coupling coefficient of .21. The coupling loop is approximately 3 mm. diameter.

Automatic Frequency Tuning Control

A digital auto tuner has been developed for the masers using medium scale integrated circuits. The system measures period changes, between the maser and a reference oscillator, produced by modulating the transition

* This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

linewidth. A proportional correction controls a cavity tuning varactor.

For tuning, the masers are offset by .01 Hz. The beat period is measured for two levels of hydrogen flux. The period difference is integrated in an updown counter, buffered, and transferred to a digital-to-analog converter. The D/A converter controls the cavity tuning varactor. Linear drifts in the tuning reference are removed by alternating the direction of the modulation and period count after each pair of counts.

A period measurement method was selected over a phase locking technique because the counter readout for 100 sec beat is a convenient system frequency monitor at 100 sec. Also, the zero point stability required of the phase detector (mixer) is more critical for a phase locking technique, i.e., the phase detector averages adjacent quarter cycles and a static zero error produces an output error. For a zero crossing detector the phase can be measured for a full cycle and, to the first order, a static error does not change the period.

The tuner will typically converge to a $\delta f/f = 2 \times 10^{-14}$ in approximately 4 hours, using a second maser as a tuning reference. For reliability, the Deep Space Network operation will use two masers at each antenna site. Also, to guarantee maser short term performance (τ's less than 1 week), a second maser is essential as a comparison reference. The tuning method to be used with two masers is to modulate one maser for cavity tuning; the other, tracking the averaged frequency of the first, will be used for the prime station frequency standard. A second digital integrator within the tuner is provided for this purpose.

Hydrogen Maser Receiver-Synthesizer Unit

The receiver-synthesizer section of the Hydrogen Maser amplifies the low level output signal of the maser to a useful power level, (+13 dbm) and produces standard output frequencies of 100 MHz, 10 MHz, 5 MHz, and 100 KHz. The receiver is a triple conversion, second order phase-locked loop with the internal voltage-controlled crystal oscillator operating at 100 MHz. The noise bandwidth of the loop was set at 100 Hz, yielding a signal-to-noise ratio of 64 db where

$$\text{Power Output of Maser} = -85 \text{ dbm}$$

$$N_o = -174 \ \frac{\text{dbm}}{\text{Hz}}$$

$$B_w = 100 \text{ Hz}$$

$$\text{Noise figure of Rcvr} = 5 \text{ db}$$

The frequency synthesizer located in the inner loop has been specially modified for this receiver. Its frequency range is 400 KHz to 510 KHz in steps of 10^{-8}Hz, thereby providing the maser with a frequency settability of 7 parts in 10^{18}. The synthesizer is driven directly from the 100 MHz distribution amplifier, thereby minimizing any spectral degradation due to additional frequency dividers. The receiver is shown in figure 3. The synthesizer has been designed for low phase drift with temperature in order to utilize its low frequency digits. The maximum rate of drift of phase for a 25°C step in temperature was 0.3×10^{-3} degrees of phase/second at 50 MHz corresponding to a frequency stability of 1.6 parts in 10^{14}.

Long Term Stability: In order not to degrade the performance of the maser, specifications were established for each RF module in the receiver concerning their short term and long term stability. The long term stability criterion for the receiver consisted of subjecting each RF module to a step change in temperature of 10°C, and recording the maximum rate of change of phase. The maximum rate of change of phase was then used to calculate the stability $\Delta\varphi$ of each module. This number, modified by any loop correction, had to be less than or equal to 5×10^{-14}.

Short Term Stability: Good short term stability requires that both amplitude and phase modulation be minimized in an amplifier. Until recently it was difficult to make low level phase and amplitude noise measurements. They were restricted to voltage controlled oscillators and frequency synthesizers where the noise was high. However, the availability of more sensitive measuring equipment has enabled more accurate measurements to be made.

Upon investigating the cause of AM and PM noise in amplifiers, it is found that if the signal power is much larger than the noise power, half the additive noise power appears as amplitude modulation sidebands and the other half as phase modulation sidebands[1]. It can be shown that if the amplitude noise and phase noise are both small, then the modulation processes can be handled separately[2]. For amplitude modulation alone define

$$A(w_1,w_2) = \frac{\text{single sideband AM power for } |w|\epsilon[w_1,w_2]}{\text{total carrier power}}$$

per device, referred to the input. Similarly for phase modulation alone define

$$P(w_1,w_2) = \frac{\text{single sideband PM power for } |w|\epsilon[w_1,w_2]}{\text{total carrier power}}$$

per device, reffered to the input.

A typical setup that is used to measure either AM or PM noise in an amplifier is shown in figure 4. To measure the PM noise, the signals at the input of the mixer are put in quadrature and the output is low pass filtered and amplified. A wave analyzer then measures the amount of power in a certain bandwidth of the spectrum. If one standardizes on a 1 Hz bandwidth at 10 Hz, then a measurement of the power or RMS voltage in this bandwidth can serve as a basis for the comparison of circuit performance. Similarly, to measure AM noise the inputs to the mixer are kept in phase and its output is coupled to the d-c amplifier with a large capacitor. The amount of AM noise will be

$$A(w_1,w_2) = \frac{V_{sb}}{V_c}_{db} = 20 \log \frac{V_N[w_1,w_2]}{2K}$$

where

$A(w_1,w_2) \triangleq$ single sided amplitude noise relative to carrier for $|w|\epsilon[w_1,w_2]$ in db.

$V_N[w_1,w_2]$ = RMS Noise voltage for $|w|\epsilon[w_1,w_2]$

K = RMS value of S-curve in volts/RAD

and the PM noise will be

$$P(w_1,w_2) = \frac{V_{sb}}{V_c}_{db} = 20 \log \frac{V_N[w_1,w_2]}{2K}$$

where

$P(w_1,w_2) \triangleq$ single sided phase noise relative to carrier for $|w|\epsilon[w_1,w_2]$

Any noise perturbation from the signal source will tend to cancel out. By using a low 1/f noise d-c amplifier and a Schottky barrier diode mixer, phase jitter in the order of 2μ degrees RMS/$\sqrt{\text{Hz}}$ could be measured at 10 dz[3]. Measurements conducted on standard frequency dividers and distribution amplifiers prompted us to design all new modules using Class A amplifiers. RF negative feedback reduced phase noise significantly in the amplifiers increasing the signal-to-phase noise ratio by 20 or 30 db.

Phase noise was not affected to any great degree by the type of transistor used, the collector current, or the 1/f characteristic of the transistor. A plot of the single sided phase noise relative to the carrier for the distribution amplifiers as well as the frequency dividers is shown in figure 5.

Frequency Stability Measurements

The frequency stability for the masers is shown in figure 6. It can be seen that the goals described above have been met. This data was originally taken when both masers were in a well-isolated environment. However, it has been duplicated with one maser in a steel-semi-trailer. The data for 10^5 sec and 10^6 sec were taken with the masers 20 km apart, via microwave link, accounting for the large variance. The short term ($\alpha\ 1/\tau$), data were not as good as expected from theoretical considerations. However, the image noise in the receiver is not filtered out in the present system. Removing this by suitable filtering will improve the short term stability by 3 db. In addition, measurement of the local oscillator frequency multiplier indicate that approximately 2-3 db improvement will be obtained when an improved frequency multiplier is installed in June. The data was taken at a maser output power of -69 dbm. Raising the power to the nominal -65 dbm will also improve the short term stability.

Two prototype units are being constructed. A total of six units will be fabricated for the DSN. One additional unit, mounted in a semi-trailer will be available for experimental work.

References

[1] W. B. Davenport and W.L. Root, Random Signals and Noise, NewwYork: McGraw-Hill, 1958, Chapter 7.

[2] G. D. Thompson, A. Sward, Analysis of Random Modulation in Amplifier Circuits, Space Programs Summary, 37-65, Vol. II, 1970.

[3] R. Meyer and A. Sward, The Measurement of Phase Jitter, Space Programs Summary 37-64, Vol. II, 1970.

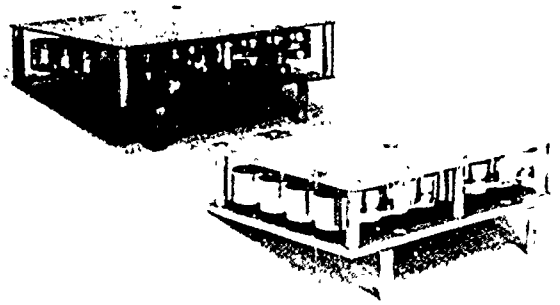


FIGURE 1 - Ion Pump Element

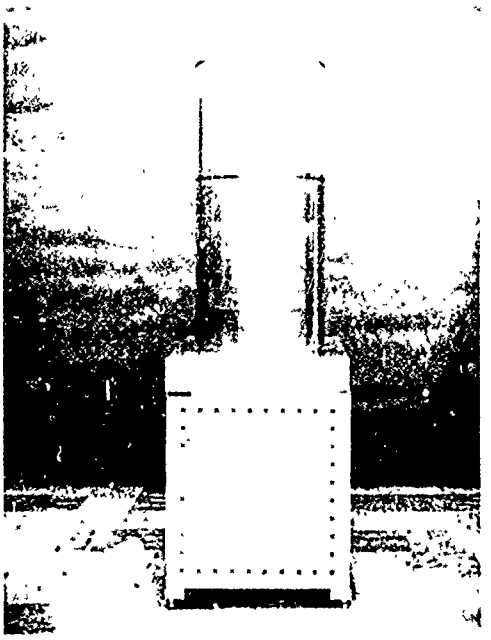


FIGURE 2 - Hydrogen Maser Physics Unit

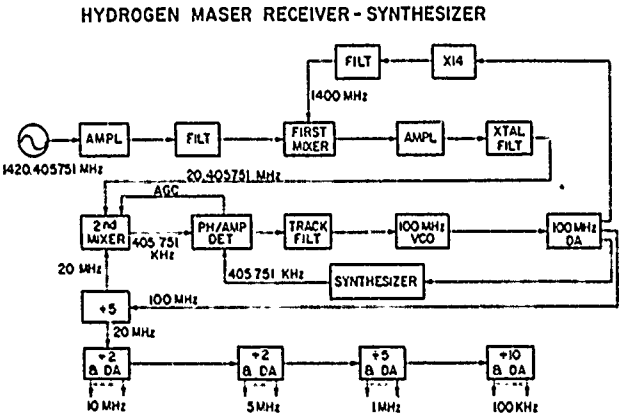


FIGURE 3

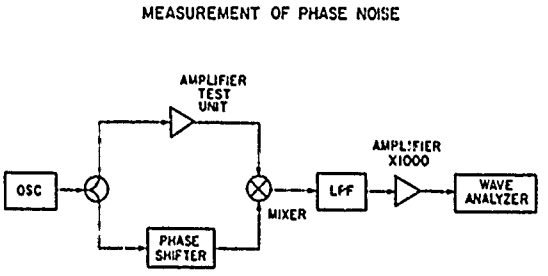


FIGURE 4

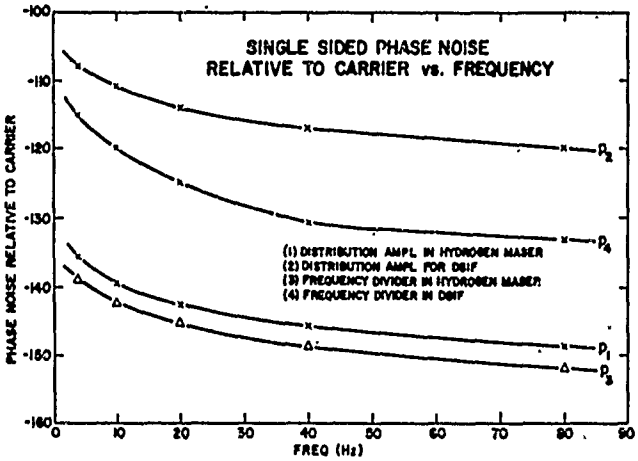


FIGURE 5

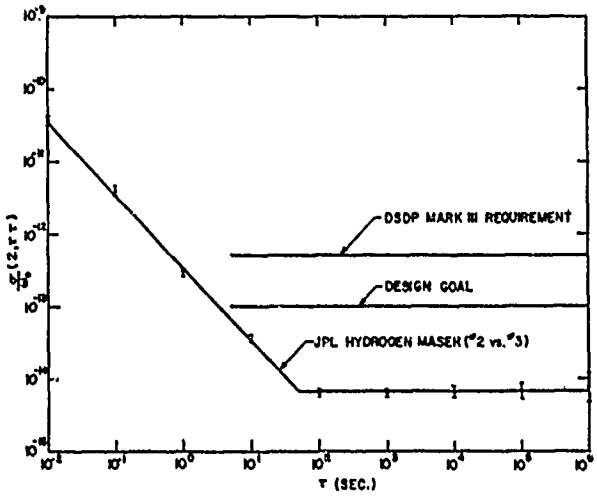


FIGURE 6